Protein Structures

A protein domain is considered to be a distinct functional and/or structural unit. A domain in a structural context refers to a segment of a polypeptide chain that can fold into an independent three-dimensional structure. It may interact with other domains of the protein or may simply be joined to other domains by a polypeptide chain. A domain in a sequence context refers to a long sequence pattern that is shared by other proteins having a common evolutionary origin. A domain may include all of the protein sequence or a part of it. A conserved domain is a recurring unit in molecular evolution whose extents can be determined by sequence and structure analysis.

The Conserved Domain Database (CDD) contains domains derived from the Smart, Pfam and Clusters of Orthologous Groups (COGs) databases. Conserved domains can be represented as multiple sequence alignments. Source alignments are processed by NCBI as follows:

- Sequences in the alignment for which a link cannot be provided to a protein in Entrez are removed.
- If possible, a closely related sequence with a known structure is substituted.
- A representative sequence, preferably with a structure link, is chosen from among those in the alignment.
- A consensus sequence is made.
- A position-specific scoring matrix (PSSM) is constructed.

The Conserved Domain search (CD-search) compares a protein sequence to the PSSMs in the CDD database to identify conserved domains within it and to identify a 3-D modeling template. Since the PSSMs are the "subject", instead of the query as in PSI-Blast, the CD-search is a form of Reverse Position-Specific Blast (RPS-Blast).

The Conserved Domain Architecture Retrieval Tool (CDART) can be used to identify proteins containing the domain(s) present in the query sequence. Conserved domain(s) present in all sequences within Entrez proteins are identified using CD-search during routine NCBI processing. These pre-computed results are accessed through CDART.

The Vector Alignment Search Tool (VAST) is a computer algorithm developed at NCBI to detect similar protein 3-dimensional structures. The "structure neighbors" for every structure in NCBI' Molecular Modeling Database (MMDB) are pre-computed. These neighbors can be used to identify distant homologs that cannot be recognized by sequence comparison alone. A VAST-search can be used for determining the structure neighbors for recently solved structures not yet in MMDB.

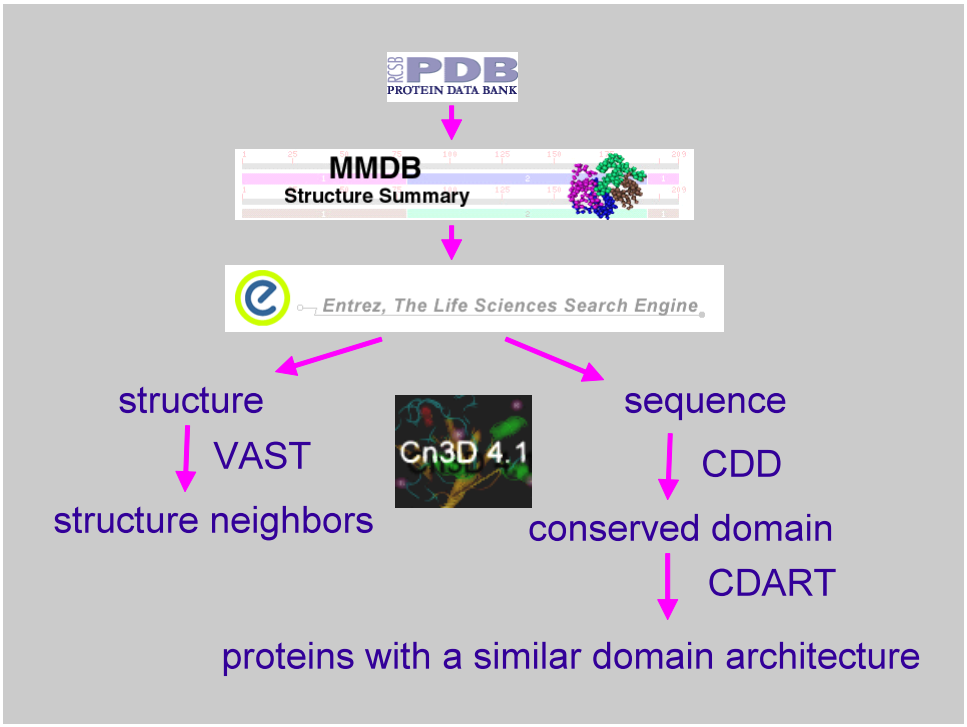
Cn3D is a helper application for web browsers to view 3-dimensional structures from NCBI's Entrez retrieval service. Cn3D runs on Windows, Macintosh, and Unix. Cn3D simultaneously displays structure, sequence, and alignment, and now has powerful annotation and alignment editing features.

In this exercise, we will learn to

- Identify a conserved domain present in the query protein using **CDD**
- Search for other proteins containing similar domain(s) using **CDART**
- Explore a 3D modeling template for the query sequence using **CDD**
- Find similar structures using **VAST**
- Visualize and annotate the 3D protein structures using **Cn3D**

The remainder of the handout includes the introductory slides and the screen shots of the exercise demonstrated in Problem 1.

Slides:



http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

Conserved Domain

- recurring unit in molecular evolution, whose extents can be determined by sequence and structure analysis
- performs a particular function
- represented as a multiple local sequence alignment of proteins containing the domain

NCBI

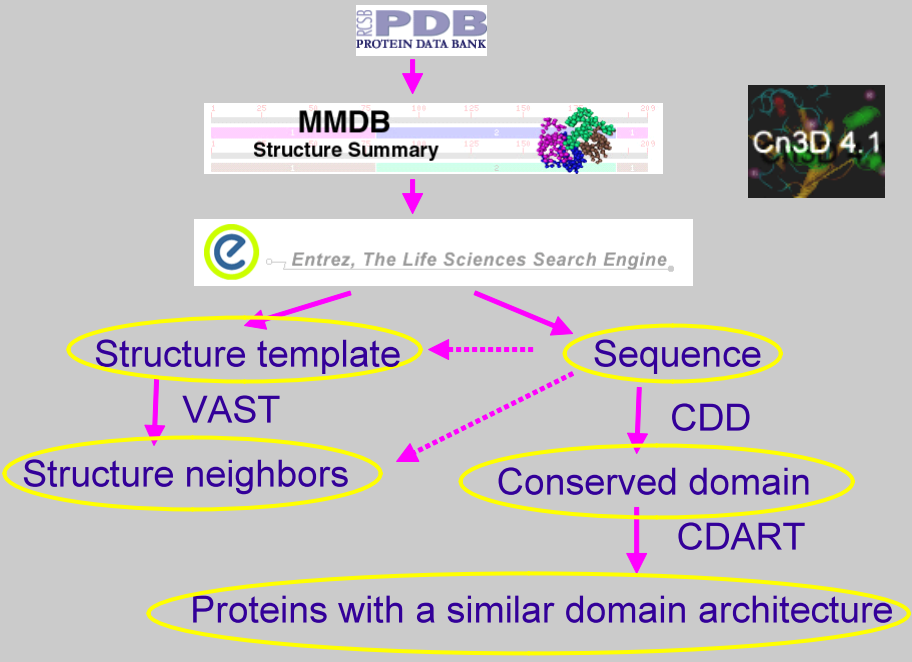
Conserved Domain Database

Pfam SMART COG KOG PRK

Curated CDs

- A position-specific scoring matrix (PSSM) is calculated
- CD-Search can be used to search against the PSSMs
- Manual curation of CDs has begun

NCBI



Exercise 1

In this problem, we will follow these steps:

A. Identify conserved domain(s) present in a protein.
B. Search for other proteins containing similar domain(s).
C. Explore a 3D modeling template for the query sequence.
D. Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence.

Navigate to the CDD Search page at:
http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

Paste the following protein sequence in the CD-Search query box and run the search.

MDPIALTAAVGADLLGDGRPETLWLGIGTLLMLIGTFYFIVKGWGVTDKEAREYY
SITILVPGIASAAYLSMFFGIGLTEVQVGSEMLDIYYARYADWLFTTPLLLLDLALL
AKVDRVSIGTLVGVDALMIVTGLVGALSHTPLARYTWWLFSTICMIVVLYFLATS
LRAAAKERGPEVASTFNTLTALVLVLWTAYPILWIIGTEGAGVVGLGIETLLFMVL
DVTAKVGFGFILLRSRAILGDTEAPEPSAGAEASAAD

A. What is the domain present in this protein?

Obtain more information about the bacteriorhodopsin domain by searching in NCBI's Bookshelf

http://www.ncbi.nlm.nih.gov/books/

B. Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domain architectures" button. To display the records, click on the link to the sequences and from there on the “Look up Sequences in Entrez”. Change the display from “Summary” to “FASTA”.

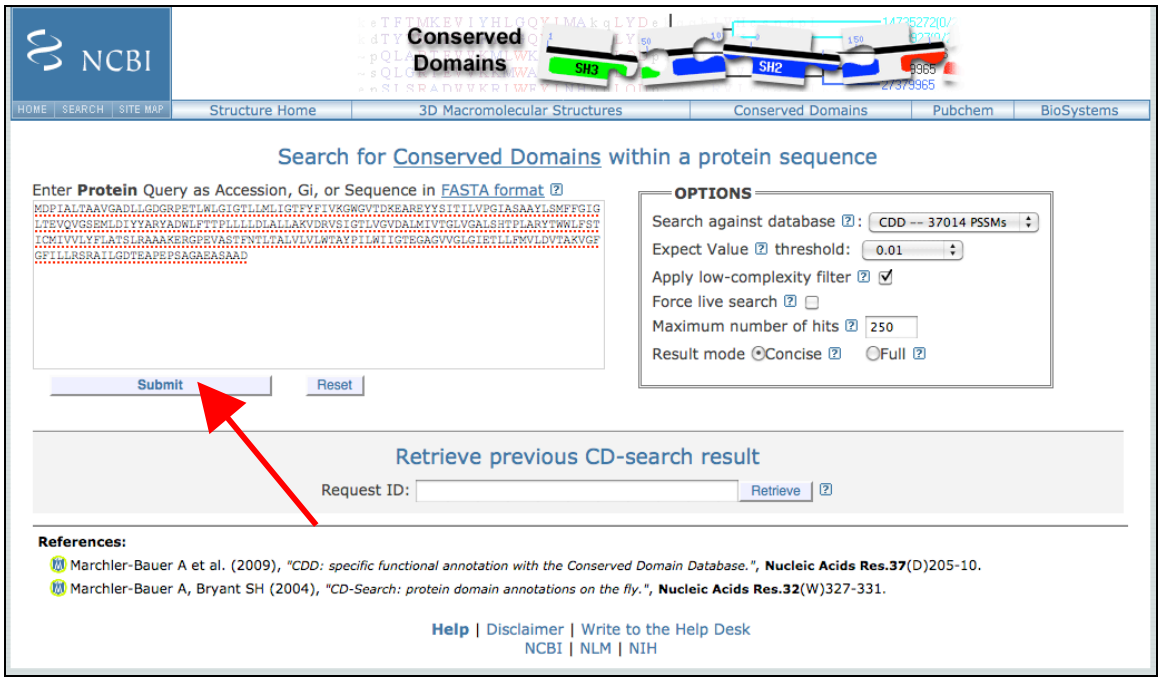
C. Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the full display button, then the Pfam domain entry graphic of the domain. Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. Extend the “Structure” display and invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Structure View" button.

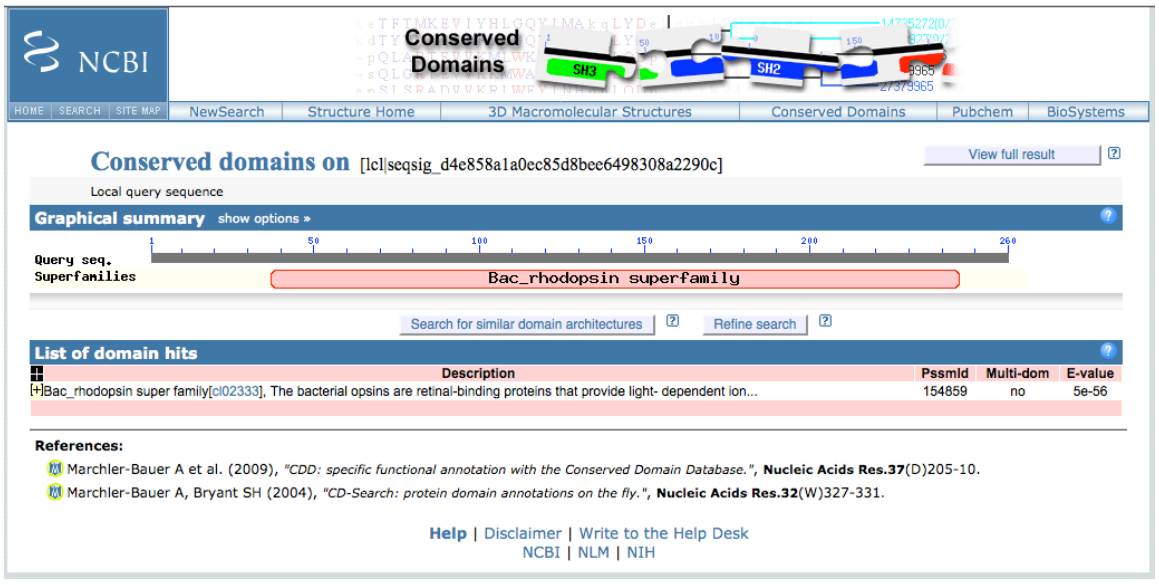
The structure of the *Halobacterium salinarum* bacteriorhodopsin mutant protein and its sequence alignment with our query protein are displayed. For a better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a bacterial rhodopsin signature (FMVLDVTAKVGF) where K is the retinal binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the halorhodopsin protein sequence.

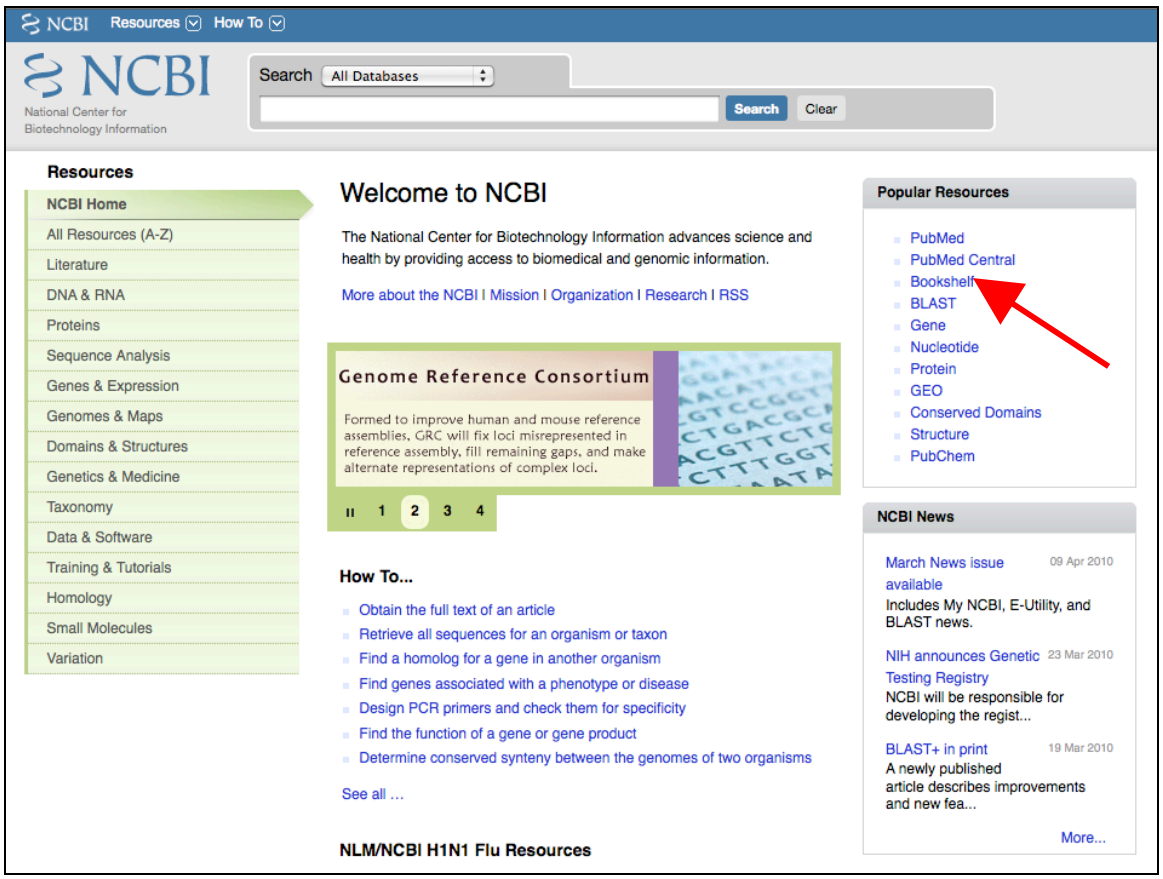
Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the “Done” button. To zoom in, press z on the keyboard. Identify the cofactor near the lysine residue.

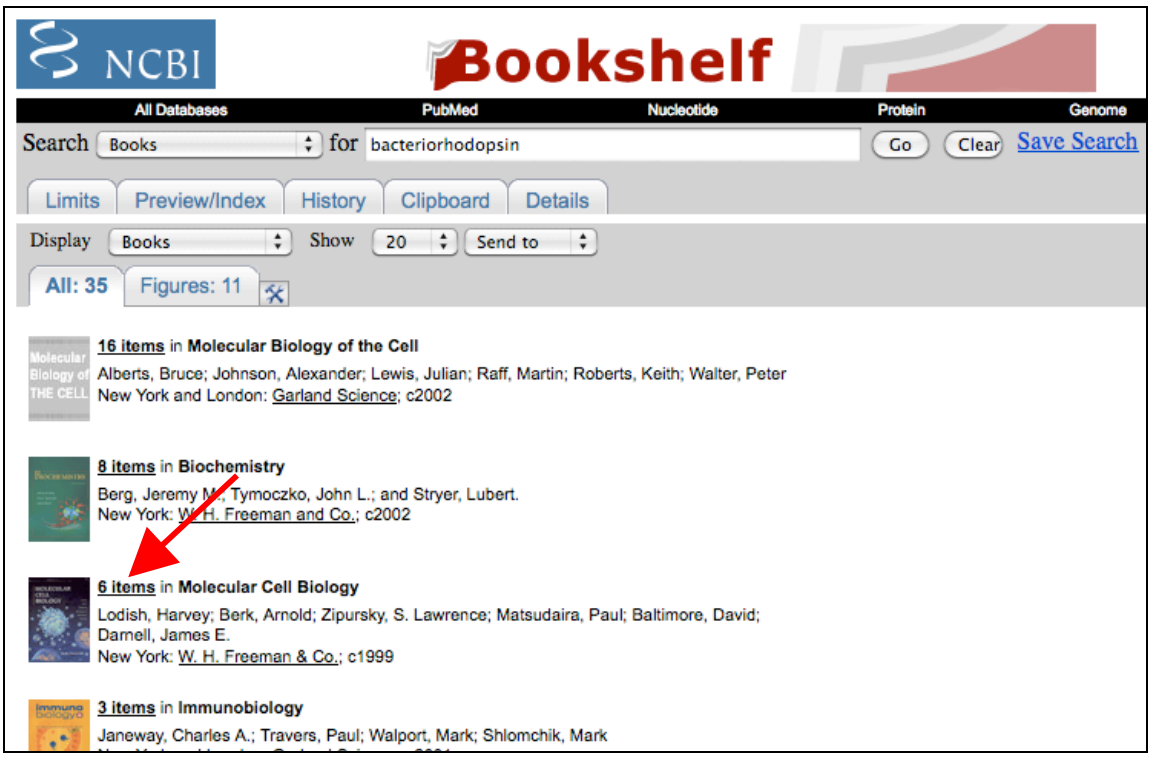
D. To obtain the structural neighbors for the halorhodopsin protein, first click on the structure entry link, 2JAF, on the CD-Browser page. Then click Links → Structure on the bottom right, then on 2JAF again in the Entrez Structure page, and finally on the grey “Sequence A” graphic. You will see a graphical representation of the regions aligned by VAST. To view the statistics in tabular format, select Table from the “in” portion of the List menu line and click list. Select one or more of the check boxes next to the structure neighbors and view by clicking on the "View 3D Structure" button.

Screen images:






NCBI Resources How To
NCBI
National Center for Biotechnology Information
Search All Databases
Search Clear
Resources
NCBI Home
All Resources (A-Z)
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes & Maps
Domains & Structures
Genetics & Medicine
Taxonomy
Data & Software
Training & Tutorials
Homology
Small Molecules
Variation
Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
More about the NCBI | Mission | Organization | Research | RSS
Genome Reference Consortium
Formed to improve human and mouse reference assemblies. GRC will fix loci misrepresented in reference assembly, fill remaining gaps, and make alternate representations of complex loci.
1 2 3 4
How To...
Obtain the full text of an article
Retrieve all sequences for an organism or taxon
Find a homolog for a gene in another organism
Find genes associated with a phenotype or disease
Design PCR primers and check them for specificity
Find the function of a gene or gene product
Determine conserved synteny between the genomes of two organisms
See all ...
NLM/NCBI H1N1 Flu Resources
Popular Resources
PubMed
PubMed Central
Bookshelf
BLAST
Gene
Nucleotide
Protein
GEO
Conserved Domains
Structure
PubChem
NCBI News
March News issue available 09 Apr 2010
Includes My NCBI, E-Utility, and BLAST news.
NIH announces Genetic Testing Registry 23 Mar 2010
NCBI will be responsible for developing the regist...
BLAST+ in print 19 Mar 2010
A newly published article describes improvements and new fea...
More...


NCBI
Bookshelf
All Databases PubMed Nucleotide Protein Genome
Search Books for bacteriorhodopsin Go Clear Save Search
Limits Preview/Index History Clipboard Details
Display Books Show 20 Send to
All: 35 Figures: 11
16 items in Molecular Biology of the Cell
Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter
New York and London: Garland Science; c2002
8 items in Biochemistry
Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert.
New York: W. H. Freeman and Co.; c2002
6 items in Molecular Cell Biology
Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E.
New York: W. H. Freeman & Co.; c1999
3 items in Immunobiology
Janeway, Charles A.; Travers, Paul; Walport, Mark; Shlomchik, Mark

Many Integral Proteins Contain Multiple Transmembrane α Helices

Although Figure 3-33 depicts glycophorin as a monomer with a single α helix spanning the bilayer, this protein is present in erythrocyte membranes as a dimer of two identical polypeptide chains. The two membrane-spanning α helices of glycophorin are thought to form a coiled-coil structure (see Figure 3-9a) stabilized by specific interactions between the amino acid side chains at the interface of the two helices. It is now known that many other transmembrane proteins contain two or more membrane-spanning α helices. For instance, the *bacterial photosynthetic reaction center (PRC)* comprises four subunits and several prosthetic groups, including four chlorophyll molecules. In this complex protein, three of the four subunits span the membrane; two of these subunits (L and M) each contain five membrane-spanning α helices (see Figure 16-40).

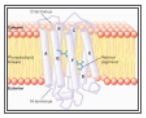


Figure 3-34
Overall structure of bacteriorhodopsin as deduced from (more...)

A large and important family of integral proteins is defined by the presence of seven membrane-spanning α helices. More than 150 such "seven-spanning" membrane proteins have been identified. This class of integral proteins is typified by *bacteriorhodopsin,* a protein found in a photosynthetic bacterium (Figure 3-34). Absorption of light by the retinal group attached to bacteriorhodopsin causes a conformational change in the protein that results in pumping of protons from the cytosol across the bacterial membrane to the extracellular space. The proton concentration gradient thus generated across the membrane is used to synthesize ATP, as discussed in Chapter 16. Both the overall arrangement of the seven α helices in bacteriorhodopsin and the identity of most of the amino acids can be resolved by computer analysis of micrographs of two-dimensional crystals of the membrane-embedded protein taken at various angles to the electron beam.

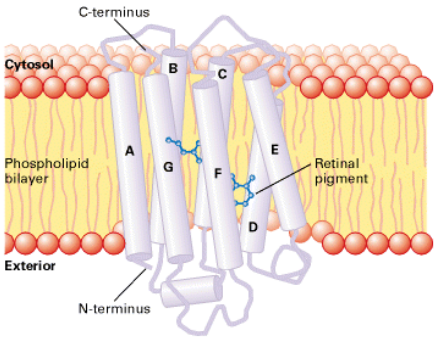
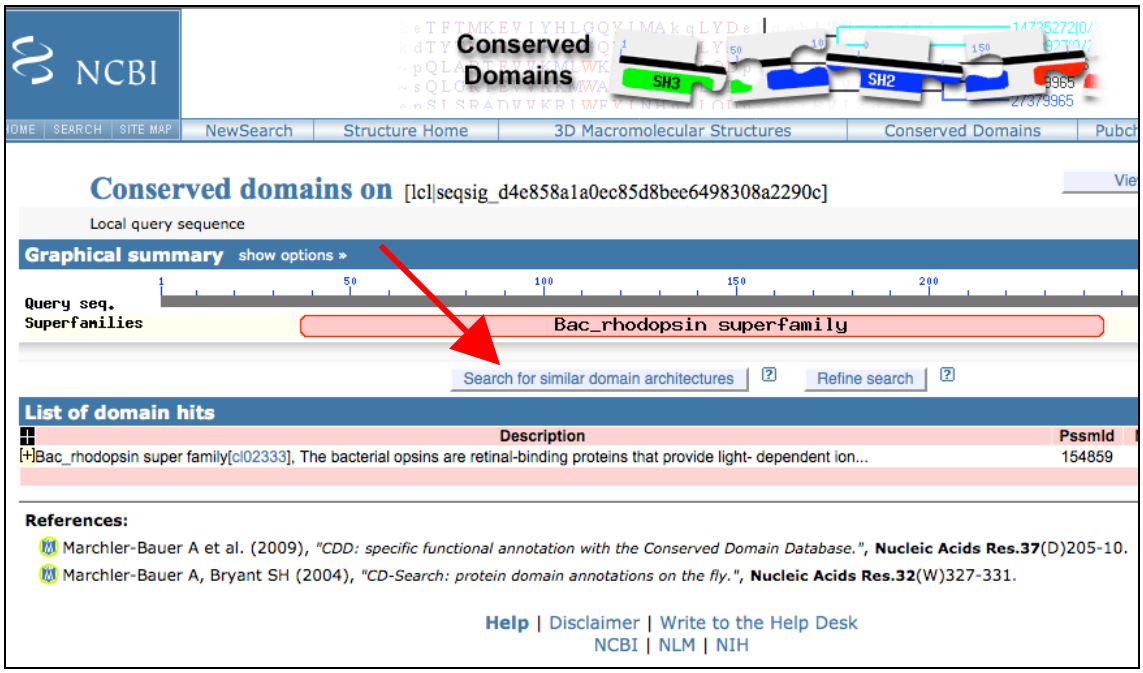
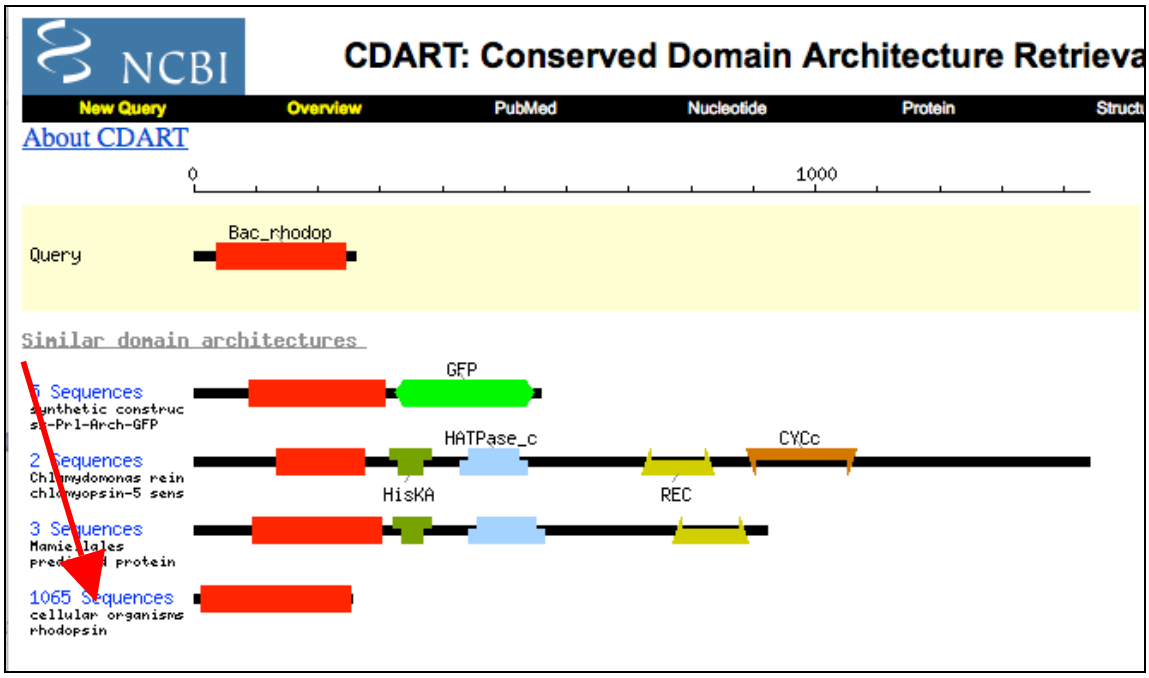
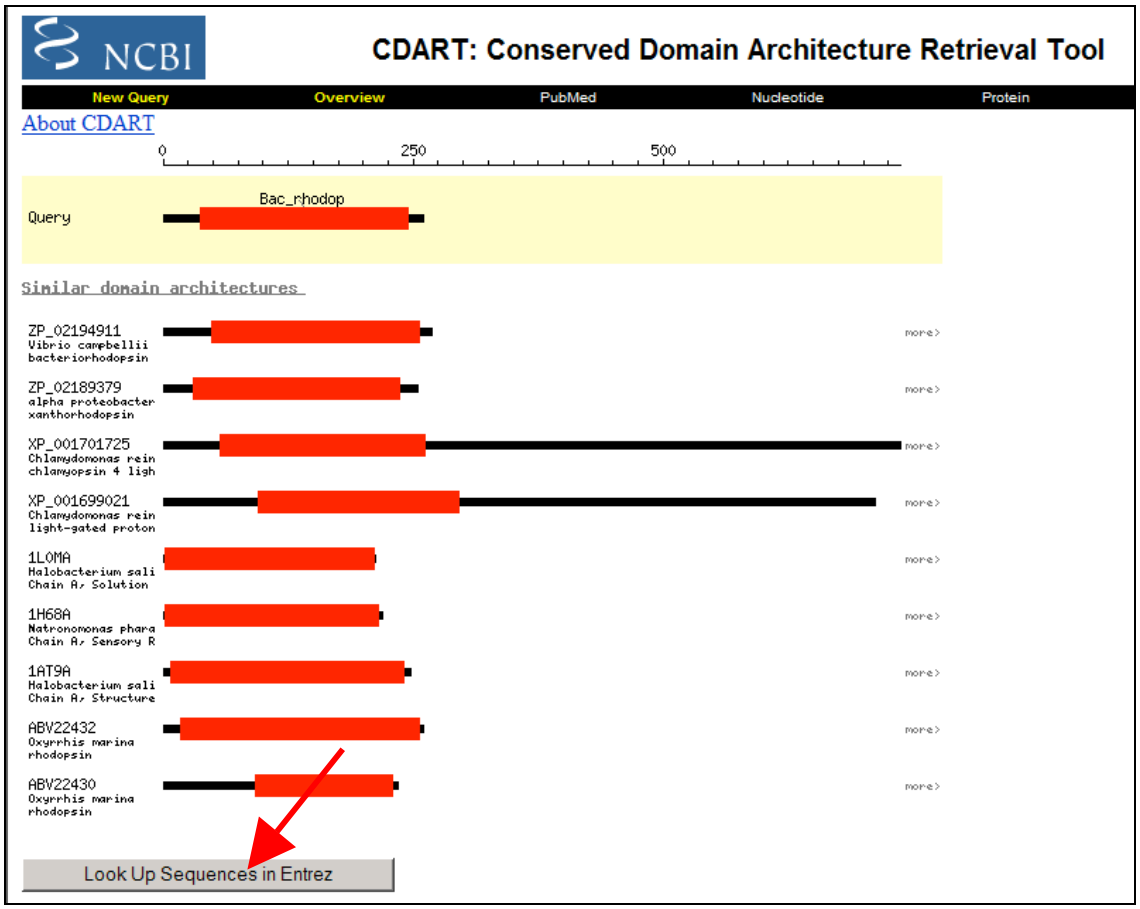


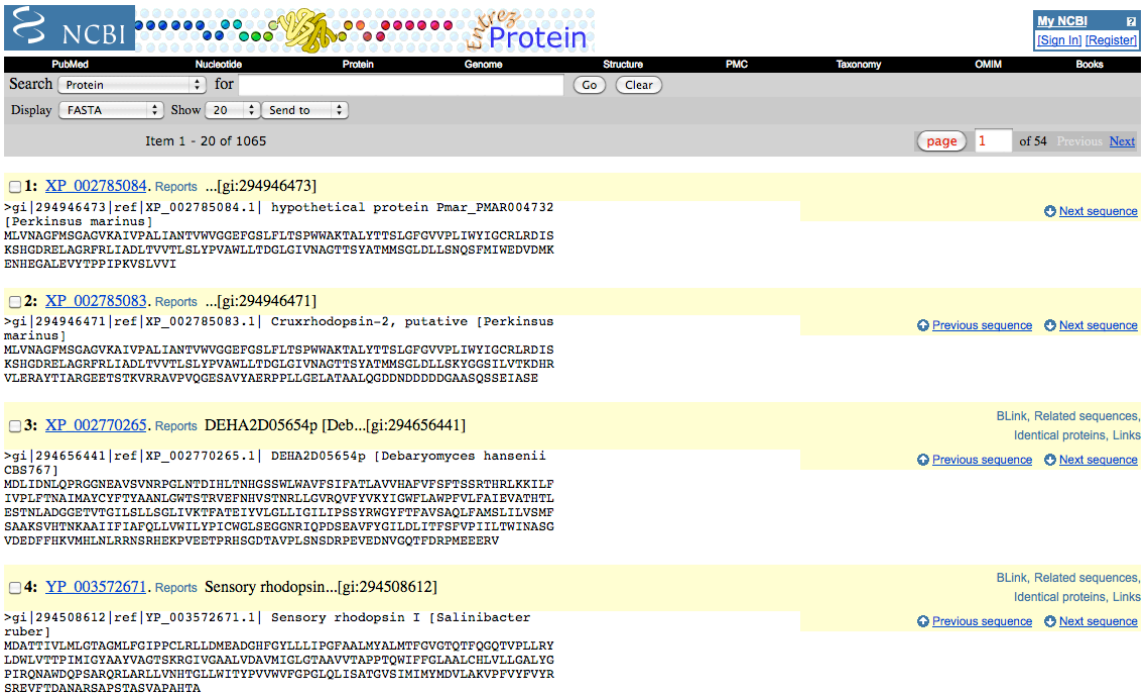
Figure 3-34. Overall structure of bacteriorhodopsin as deduced from electron diffraction analyses of two-dimensional crystals of the protein in the bacterial membrane. The seven membrane-spanning α helices are labeled A–G. The retinal pigment is covalently attached to lysine 216 in helix G. The approximate position of the protein in the phospholipid bilayer is indicated. [Adapted from R. Henderson et al., 1990, *J. Mol. Biol.* **213**:899.

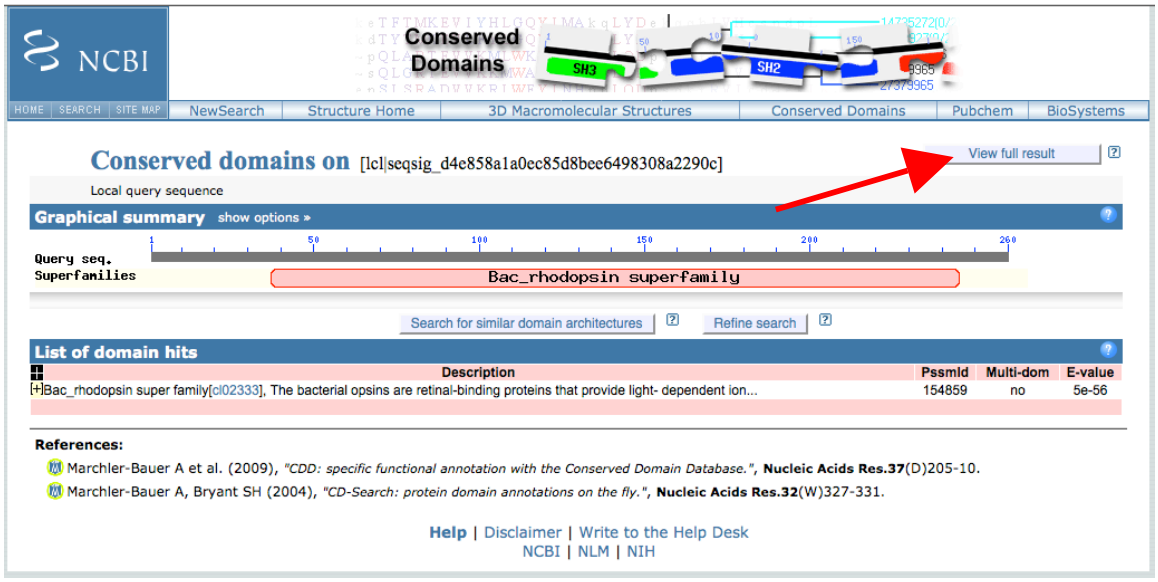
Navigation
About this book
3. Protein Structure and Function
3.1. Hierarchical Structure of Proteins
3.2. Folding, Modification, and Degradation of Proteins
3.3. Functional Design of Proteins
3.4. Membrane Proteins
3.5. Purifying, Detecting, and Characterizing Proteins
PERSPECTIVES for the Future
PERSPECTIVES in the Literature
Testing Yourself on the Concepts
MCAT/GRE-Style Questions
References

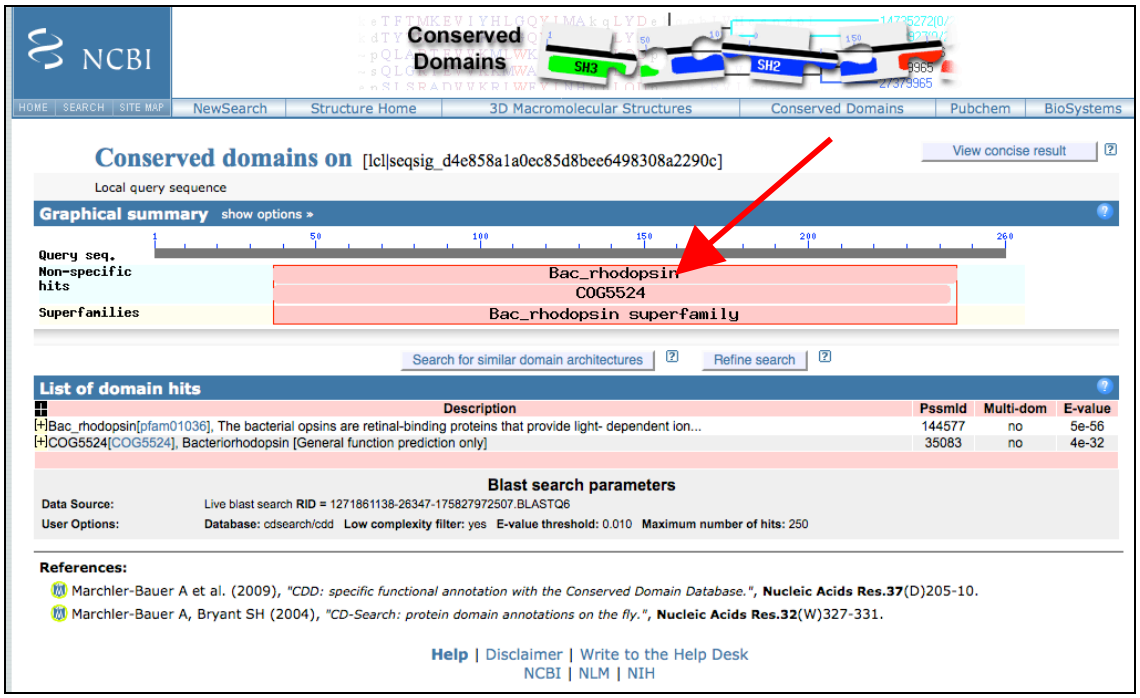

NCBI
Conserved Domains
SH3
SH2
HOME SEARCH SITE MAP
NewSearch
Structure Home
3D Macromolecular Structures
Conserved Domains
Conserved domains on [lcl|seqsig_d4e858a1a0ec85d8bee6498308a2290c]
Local query sequence
Graphical summary show options »
Query seq.
Superfamilies
Bac_rhodopsin superfamily
Search for similar domain architectures
Refine search
List of domain hits
Description
PssmId
[+]Bac_rhodopsin super family[cl02333], The bacterial opsins are retinal-binding proteins that provide light- dependent ion...
154859
References:
Marchler-Bauer A et al. (2009), "CDD: specific functional annotation with the Conserved Domain Database.", Nucleic Acids Res.37(D)205-10.
Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", Nucleic Acids Res.32(W)327-331.
Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

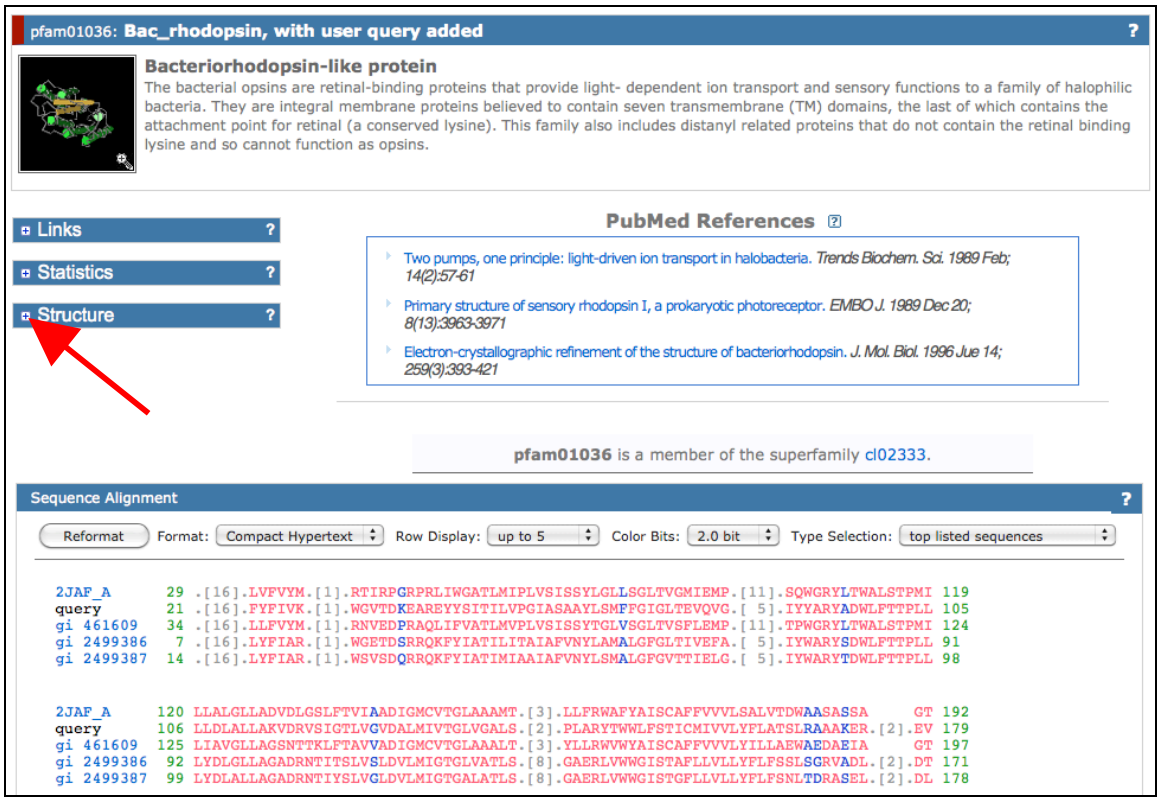

NCBI
CDART: Conserved Domain Architecture Retrieva
New Query
Overview
PubMed
Nucleotide
Protein
About CDART
0
1000
Query
Bac_rhodop
Similar domain architectures
GFP
HATPase_c
CYCc
HisKA
REC
1065 Sequences
cellular organisms
rhodopsin

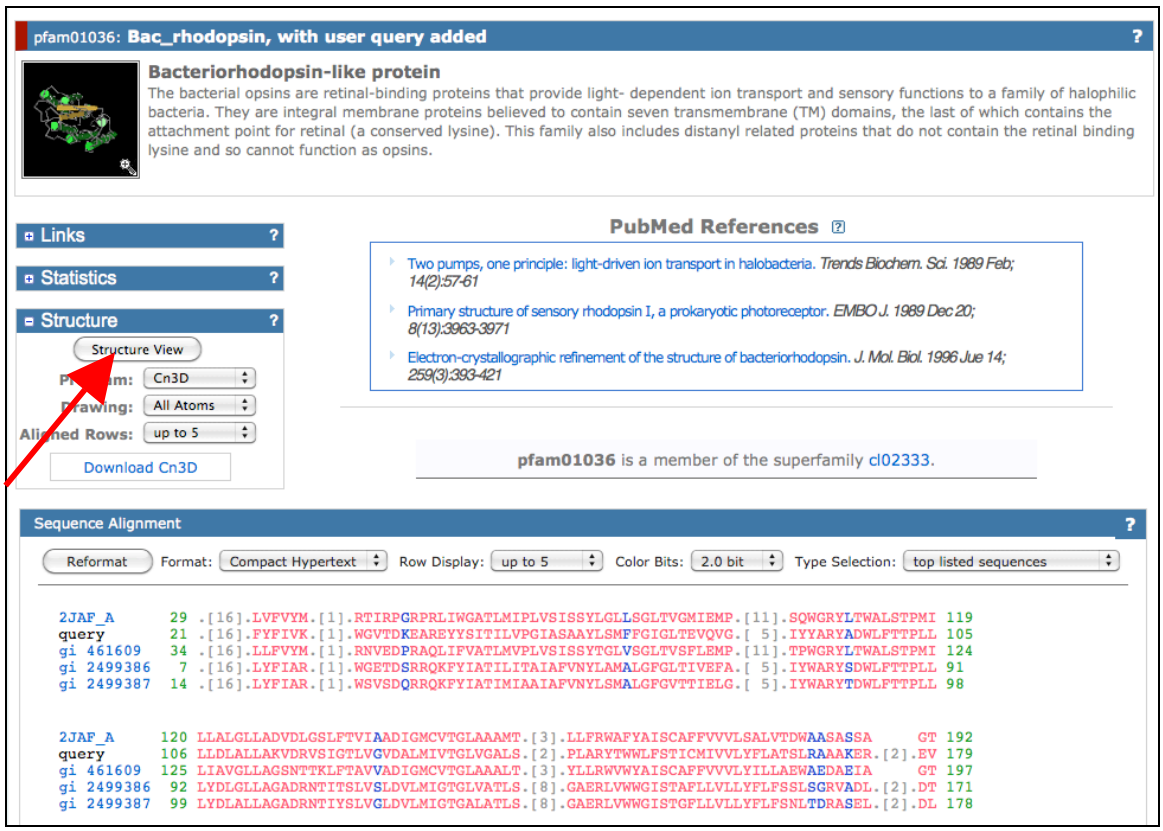

NCBI
CDART: Conserved Domain Architecture Retrieval Tool
New Query
Overview
PubMed
Nucleotide
Protein
About CDART
0
250
500
Query
Bac_rhodop
Similar domain architectures
ZP_02194911
Vibrio campbellii bacteriorhodopsin
ZP_02189379
alpha proteobacter xanthorhodopsin
XP_001701725
Chlamydomonas rein chlamyopsin 4 ligh
XP_001699021
Chlamydomonas rein light-gated proton
1L0MA
Halobacterium sali Chain A, Solution
1H68A
Natronomonas phara Chain A, Sensory R
1AT9A
Halobacterium sali Chain A, Structure
ABV22432
Oxyrrhis marina rhodopsin
ABV22430
Oxyrrhis marina rhodopsin
more>
Look Up Sequences in Entrez

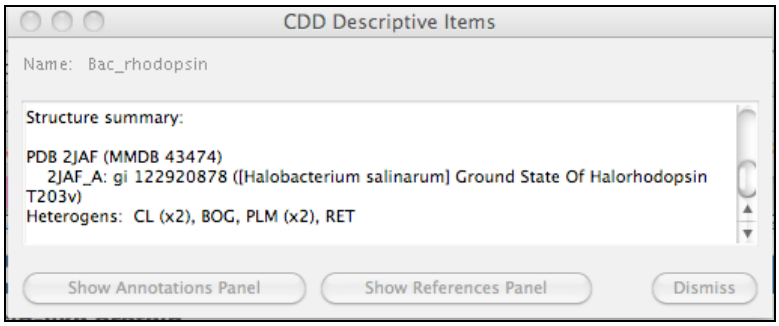

NCBI
Entrez Protein
My NCBI
[Sign In] [Register]
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
OMIM
Books
Search Protein for
Go
Clear
Display FASTA
Show 20
Send to
Item 1 - 20 of 1065
page 1 of 54 Previous Next
1: XP_002785084. Reports ...[gi:294946473]
>gi|294946473|ref|XP_002785084.1| hypothetical protein Pmar_PMAR004732 [Perkinsus marinus]
MLVNAGFMSGAGVKAIVPALIANTVWVGGEFGSLFLTSPWWAKTALYTTSLGFGVVPLIWYIGCRLRDIS
KSHGDRELAGRFRLIADLTVVTLSLYPVAWLLTDGLGIVNAGTTSYATMMSGLDLLSNQSFMIWEDVDMK
ENHEGALEVYTPPIPKVSLVVI
Next sequence
2: XP_002785083. Reports ...[gi:294946471]
>gi|294946471|ref|XP_002785083.1| Cruxrhodopsin-2, putative [Perkinsus marinus]
MLVNAGFMSGAGVKAIVPALIANTVWVGGEFGSLFLTSPWWAKTALYTTSLGFGVVPLIWYIGCRLRDIS
KSHGDRELAGRFRLIADLTVVTLSLYPVAWLLTDGLGIVNAGTTSYATMMSGLDLLSKYGGSILVTKDHR
VLERAYTIARGEETSTKVRRAVPVQGESAVYAERPPLLGELATAALQGDDNDDDDDGAASQSSEIASE
Previous sequence
Next sequence
3: XP_002770265. Reports DEHA2D05654p [Deb...[gi:294656441]
BLink, Related sequences, Identical proteins, Links
>gi|294656441|ref|XP_002770265.1| DEHA2D05654p [Debaryomyces hansenii CBS767]
MDLIDNLQPRGGNEAVSVNRPGLNTDIHLTNHGSSWLWAVFSIFATLAVVHAPVFSFTSSRTHRLKKILF
IVPLFTNAIMAYCYFTYAANLGWTSTRVEFNHVSTNRLLGVRQVFYVKYIGWFLAWPFVLFAIEVATHTL
ESTNLADGGETVTGILSLLSGLIVKTFATEIYVLGLLIGILIPSSYRWGYFTFAVSAQLFAMSLILVSMF
SAAKSVHTNKAAIIFIAFQLLVWILYPICWGLSEGGNRIQPDSEAVFYGILDLITFSPVPIILTWINASG
VDEDFFHKVMHLNLRRNSRHEKPVEETPRHSGDTAVPLSNSDRPEVEDNVGQTFDRPMEEERV
BLink, Related sequences, Identical proteins, Links
4: YP_003572671. Reports Sensory rhodopsin...[gi:294508612]
>gi|294508612|ref|YP_003572671.1| Sensory rhodopsin I [Salinibacter ruber]
MDATTIVLMLGTAGMLPGIPPCLRLLDMEADGHFGYLLLIPGFAALMYALMTFGVGTQTFQGQTVPLLRY
LDWLVTTPIMIGYAAYVAGTSKRGIVGAALVDAVMIGLGTAAVVTAPPTQWIFFGLAALCHLVLLGALYG
PIRQNAWDQPSARQRLARLLVNHTGLLWITYPVVWVFGPGLQLISATGVSIMIMYMDVLAKVPFVYFVYR
SREVFTDANARSAPSTASVAPAHTA
Previous sequence
Next sequence

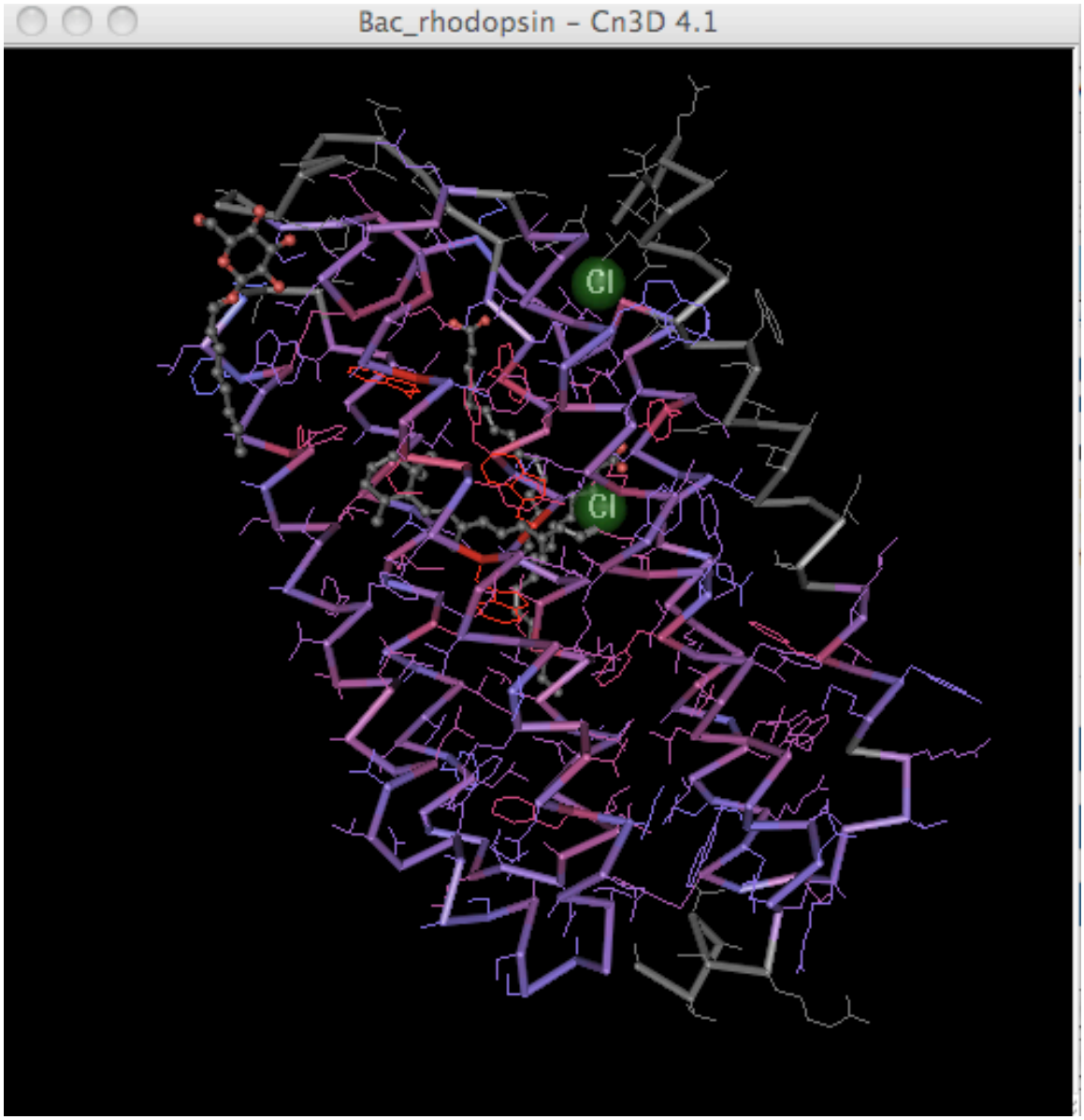

NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
NewSearch
Structure Home
3D Macromolecular Structures
Conserved Domains
Pubchem
BioSystems
Conserved domains on [lcl|seqsig_d4e858a1a0ec85d8bee6498308a2290c]
View full result
Local query sequence
Graphical summary show options »
Query seq.
Superfamilies
Bac_rhodopsin superfamily
Search for similar domain architectures
Refine search
List of domain hits
Description
Pssmid
Multi-dom
E-value
[+]Bac_rhodopsin super family[cl02333], The bacterial opsins are retinal-binding proteins that provide light- dependent ion...
154859
no
5e-56
References:
Marchler-Bauer A et al. (2009), "CDD: specific functional annotation with the Conserved Domain Database.", Nucleic Acids Res.37(D)205-10.
Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", Nucleic Acids Res.32(W)327-331.
Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

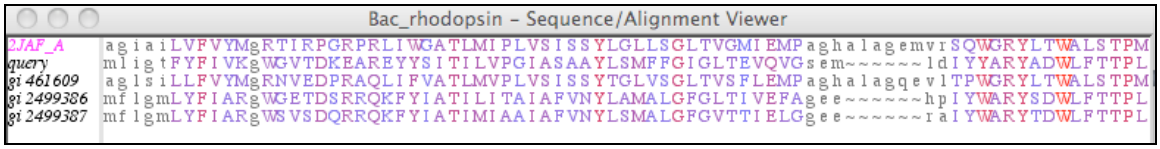

NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
NewSearch
Structure Home
3D Macromolecular Structures
Conserved Domains
Pubchem
BioSystems
Conserved domains on [lcl|seqsig_d4e858a1a0ec85d8bee6498308a2290c]
View concise result
Local query sequence
Graphical summary show options »
Query seq.
Non-specific hits
Bac_rhodopsin
COG5524
Superfamilies
Bac_rhodopsin superfamily
Search for similar domain architectures
Refine search
List of domain hits
Description
Pssmid
Multi-dom
E-value
[+]Bac_rhodopsin[pfam01036], The bacterial opsins are retinal-binding proteins that provide light- dependent ion...
144577
no
5e-56
[+]COG5524[COG5524], Bacteriorhodopsin [General function prediction only]
35083
no
4e-32
Blast search parameters
Data Source:
Live blast search RID = 1271861138-26347-175827972507.BLASTQ6
User Options:
Database: cdsearch/cdd Low complexity filter: yes E-value threshold: 0.010 Maximum number of hits: 250
References:
Marchler-Bauer A et al. (2009), "CDD: specific functional annotation with the Conserved Domain Database.", Nucleic Acids Res.37(D)205-10.
Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", Nucleic Acids Res.32(W)327-331.
Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

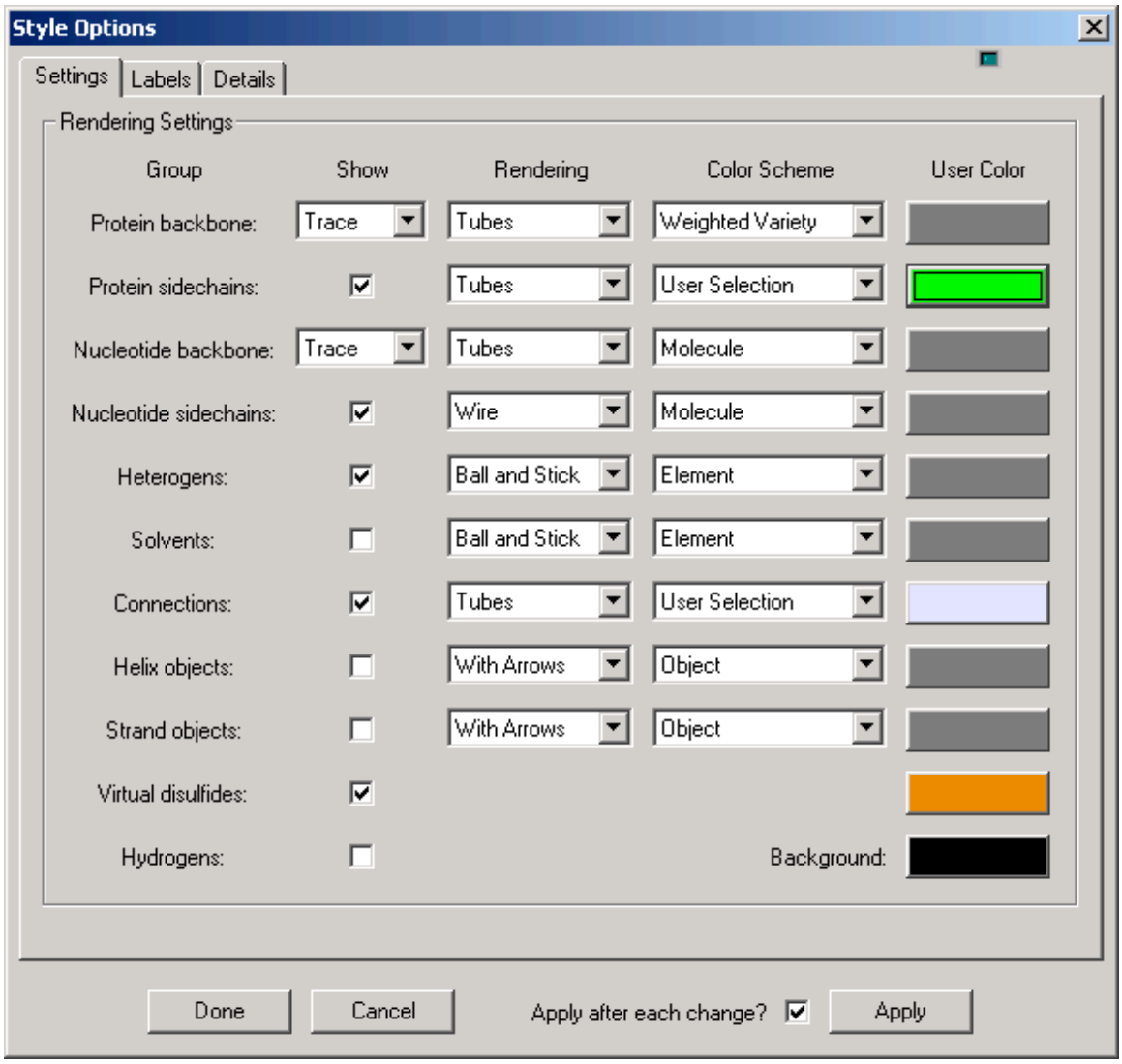

pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin-like protein
The bacterial opsins are retinal-binding proteins that provide light- dependent ion transport and sensory functions to a family of halophilic bacteria. They are integral membrane proteins believed to contain seven transmembrane (TM) domains, the last of which contains the attachment point for retinal (a conserved lysine). This family also includes distanyl related proteins that do not contain the retinal binding lysine and so cannot function as opsins.
Links
Statistics
Structure
PubMed References
Two pumps, one principle: light-driven ion transport in halobacteria. Trends Biochem. Sci. 1989 Feb; 14(2):57-61
Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor. EMBO J. 1989 Dec 20; 8(13):3963-3971
Electron-crystallographic refinement of the structure of bacteriorhodopsin. J. Mol. Biol. 1996 Jue 14; 259(3):393-421
pfam01036 is a member of the superfamily cl02333.
Sequence Alignment
Reformat
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bit
Type Selection: top listed sequences
2JAF_A 29 .[16].LVFVYM.[1].RTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGMIEMP.[11].SQWGRYLTWALSTPMI 119
query 21 .[16].LVFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[5].IYYARYADWLFTTPLL 105
gi 461609 34 .[16].LLFVYM.[1].RNVEDPRAQLIFVATLMVPLVSISSYTGLVSGLTVSFLEMP.[11].TPWGRYLTWALSTPMI 124
gi 2499386 7 .[16].LYFIAR.[1].WGETDSRRQKFYIATILITAIAFVNYLAMALGFGLTIVEFA.[5].IYWARYSDWLFTTPLL 91
gi 2499387 14 .[16].LYFIAR.[1].WSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[5].IYWARYTDWLFTTPLL 98
2JAF_A 120 LLALGLLADVDLGSLFTVIAADIGMCVTGLAAAMT.[3].LLFRWAFYAISCAFFVVVLSALVTDWAASASSA GT 192
query 106 LLDLALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[2].PLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EV 179
gi 461609 125 LIAVGLLAGSNTTKLFTAVVADIGMCVTGLAAALT.[3].YLLRWVWYAISCAFFVVVLYILLAEWAEDAEIA GT 197
gi 2499386 92 LYDLGLLAGADRNTITSLVSLDVLMIGTGLVATLS.[8].GAERLVWWGISTAFLLVLLYFLFSSLSGRVADL.[2].DT 171
gi 2499387 99 LYDLALLAGADRNTIYSLVGLDVLMIGTGALATLS.[8].GAERLVWWGISTGFLLVLLYFLFSNLTDRASEL.[2].DL 178

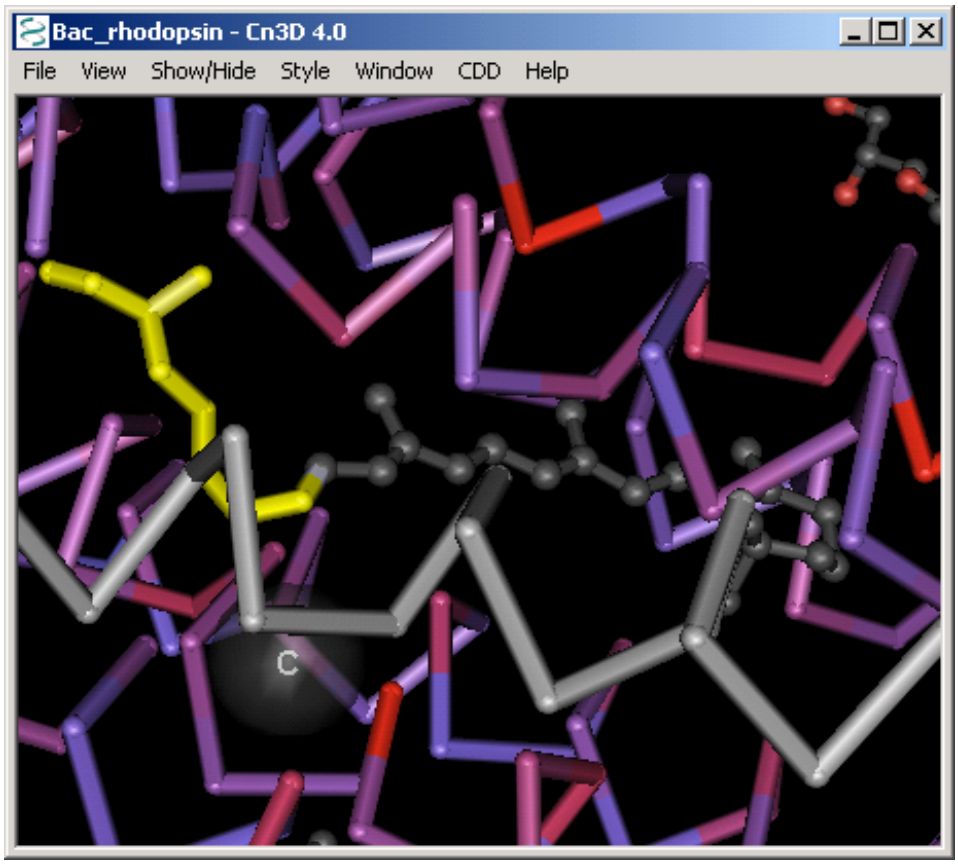

pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin-like protein
The bacterial opsins are retinal-binding proteins that provide light- dependent ion transport and sensory functions to a family of halophilic bacteria. They are integral membrane proteins believed to contain seven transmembrane (TM) domains, the last of which contains the attachment point for retinal (a conserved lysine). This family also includes distanyl related proteins that do not contain the retinal binding lysine and so cannot function as opsins.
Links
Statistics
Structure
Structure View
Cn3D
All Atoms
up to 5
Download Cn3D
PubMed References
Two pumps, one principle: light-driven ion transport in halobacteria. Trends Biochem. Sci. 1989 Feb; 14(2):57-61
Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor. EMBO J. 1989 Dec 20; 8(13):3963-3971
Electron-crystallographic refinement of the structure of bacteriorhodopsin. J. Mol. Biol. 1996 Jue 14; 259(3):393-421
pfam01036 is a member of the superfamily cl02333.
Sequence Alignment
Reformat
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bit
Type Selection: top listed sequences
2JAF_A 29 .[16].LVFVYM.[1].RTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGMIEMP.[11].SQWGRYLTWALSTPMI 119
query 21 .[16].LVFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[5].IYYARYADWLFTTPLL 105
gi 461609 34 .[16].LLFVYM.[1].RNVEDPRAQLIFVATLMVPLVSISSYTGLVSGLTVSFLEMP.[11].TPWGRYLTWALSTPMI 124
gi 2499386 7 .[16].LYFIAR.[1].WGETDSRRQKFYIATILITAIAFVNYLAMALGFGLTIVEFA.[5].IYWARYSDWLFTTPLL 91
gi 2499387 14 .[16].LYFIAR.[1].WSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[5].IYWARYTDWLFTTPLL 98
2JAF_A 120 LLALGLLADVDLGSLFTVIAADIGMCVTGLAAAMT.[3].LLFRWAFYAISCAFFVVVLSALVTDWAASASSA GT 192
query 106 LLDLALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[2].PLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EV 179
gi 461609 125 LIAVGLLAGSNTTKLFTAVVADIGMCVTGLAAALT.[3].YLLRWVWYAISCAFFVVVLYILLAEWAEDAEIA GT 197
gi 2499386 92 LYDLGLLAGADRNTITSLVSLDVLMIGTGLVATLS.[8].GAERLVWWGISTAFLLVLLYFLFSSLSGRVADL.[2].DT 171
gi 2499387 99 LYDLALLAGADRNTIYSLVGLDVLMIGTGALATLS.[8].GAERLVWWGISTGFLLVLLYFLFSNLTDRASEL.[2].DL 178

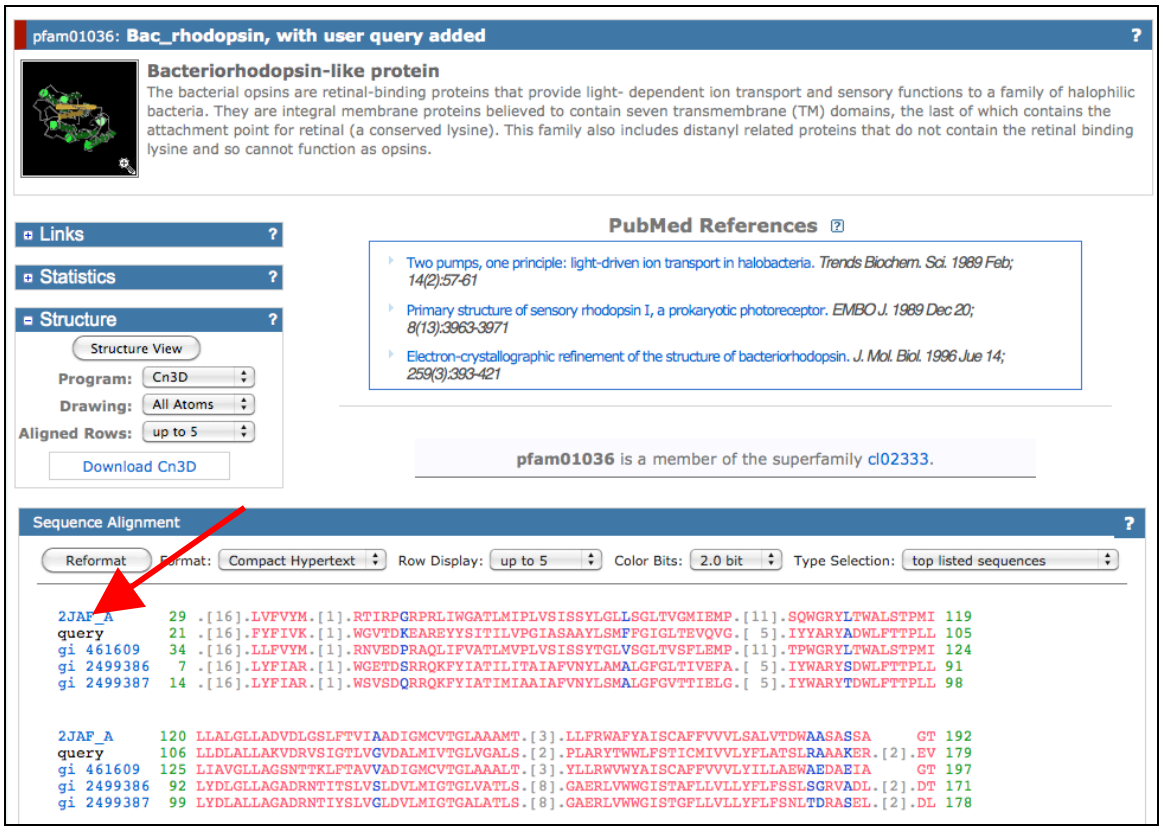

CDD Descriptive Items
Name: Bac_rhodopsin
Structure summary:
PDB 2JAF (MMDB 43474)
2JAF_A: gi 122920878 ([Halobacterium salinarum] Ground State Of Halorhodopsin T203v)
Heterogens: CL (x2), BOG, PLM (x2), RET
Show Annotations Panel
Show References Panel
Dismiss

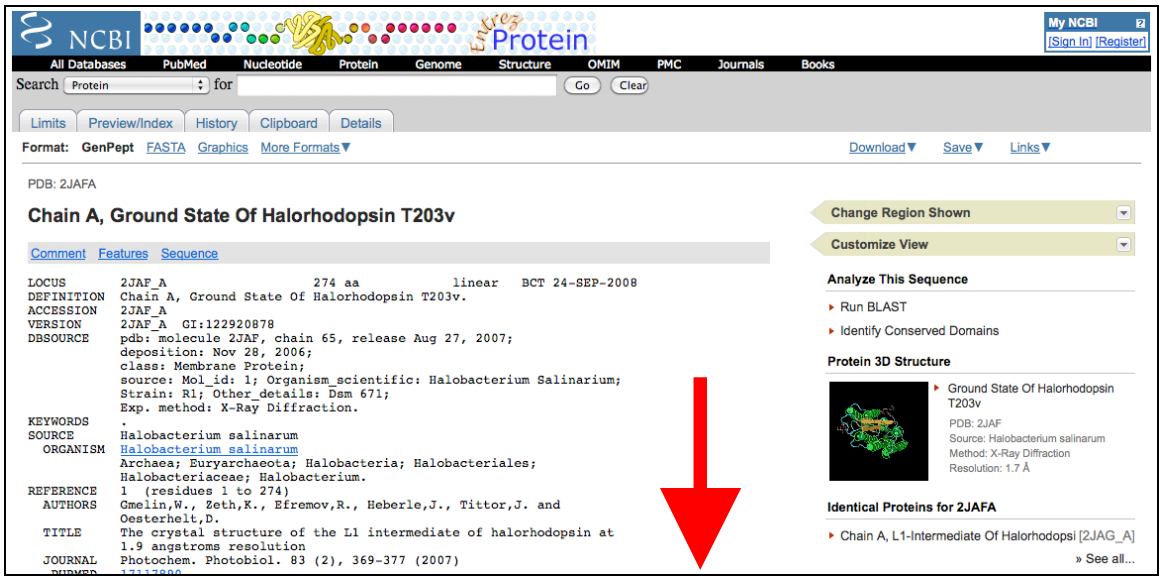

Bac_rhodopsin – Cn3D 4.1
Cl
Cl

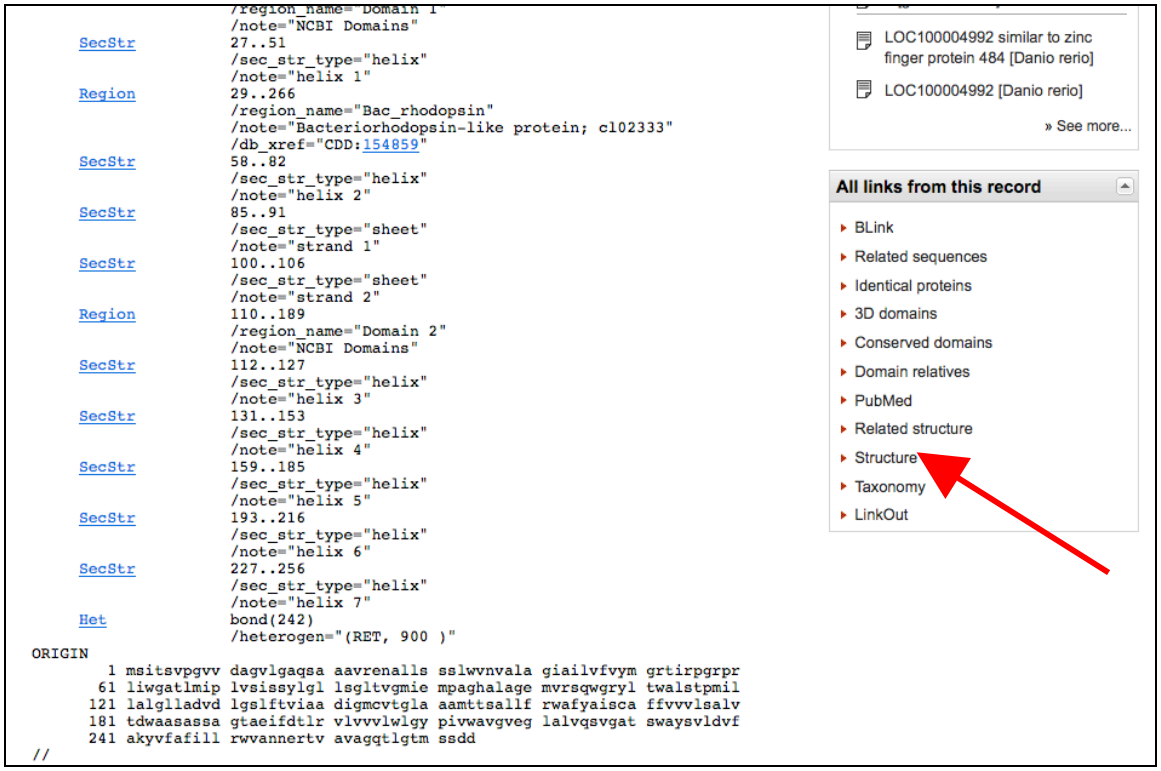

Bac_rhodopsin – Sequence/Alignment Viewer
2JAF_A agiaiLVFVYMgRTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGMIEMPaghalagemvrSQWGRYLTWALSTPM
query mligtFYFIVKgWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGsem~~~~~~ldIYYARYADWLFTTPL
gi 461609 aglsiLLFVYMgRNVEDPRAQLIFVATLMVPLVSISSYTGLVSGLTVSFLEMPaghalagqevltPWGRYLTWALSTPM
gi 2499386 mflgmLYFIARgWGETDSRRQKFYIATILITAIAFVNYLAMALGFGLTIVEFAgee~~~~~~hpIYWARYSDWLFTTPL
gi 2499387 mflgmLYFIARgWSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELGgee~~~~~~raIYWARYTDWLFTTPL

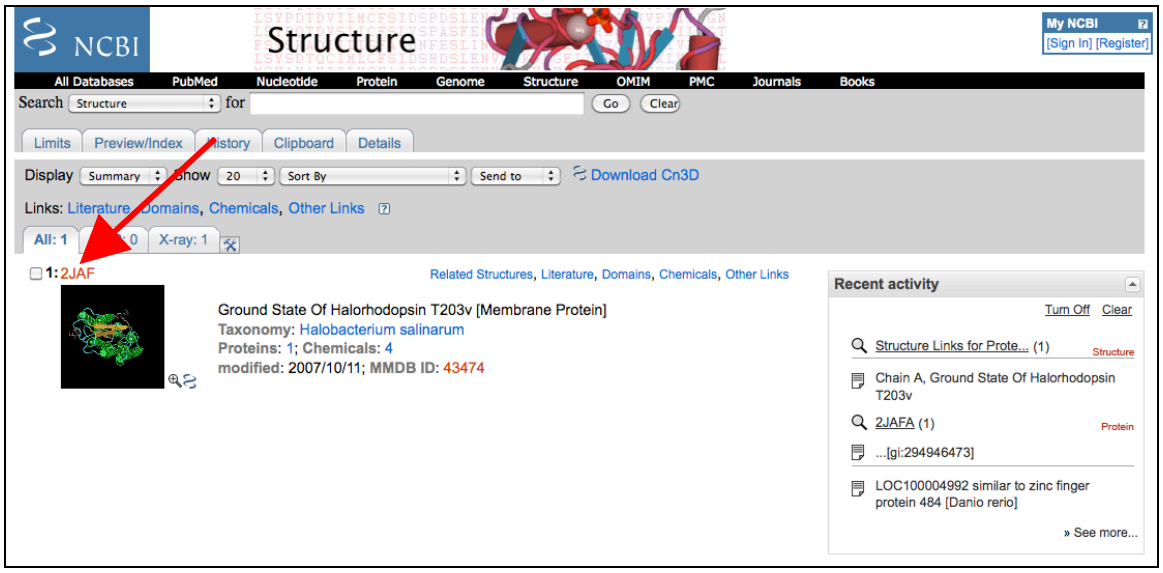

Style Options
Settings
Labels
Details
Rendering Settings
Group
Show
Rendering
Color Scheme
User Color
Protein backbone:
Trace
Tubes
Weighted Variety
Protein sidechains:
Tubes
User Selection
Nucleotide backbone:
Trace
Tubes
Molecule
Nucleotide sidechains:
Wire
Molecule
Heterogens:
Ball and Stick
Element
Solvents:
Ball and Stick
Element
Connections:
Tubes
User Selection
Helix objects:
With Arrows
Object
Strand objects:
With Arrows
Object
Virtual disulfides:
Hydrogens:
Background:
Done
Cancel
Apply after each change?
Apply

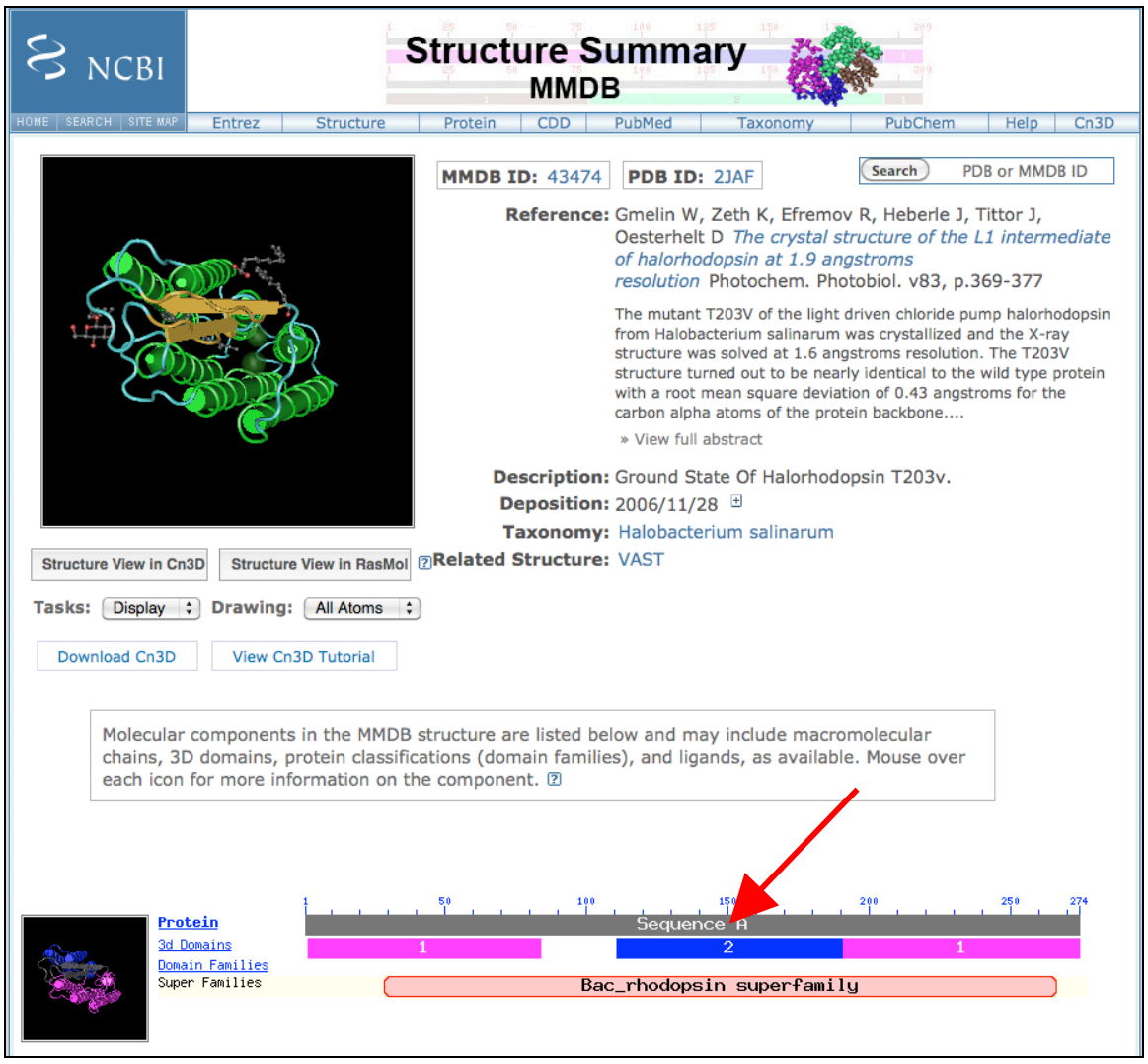

Bac_rhodopsin - Cn3D 4.0
File View Show/Hide Style Window CDD Help
C

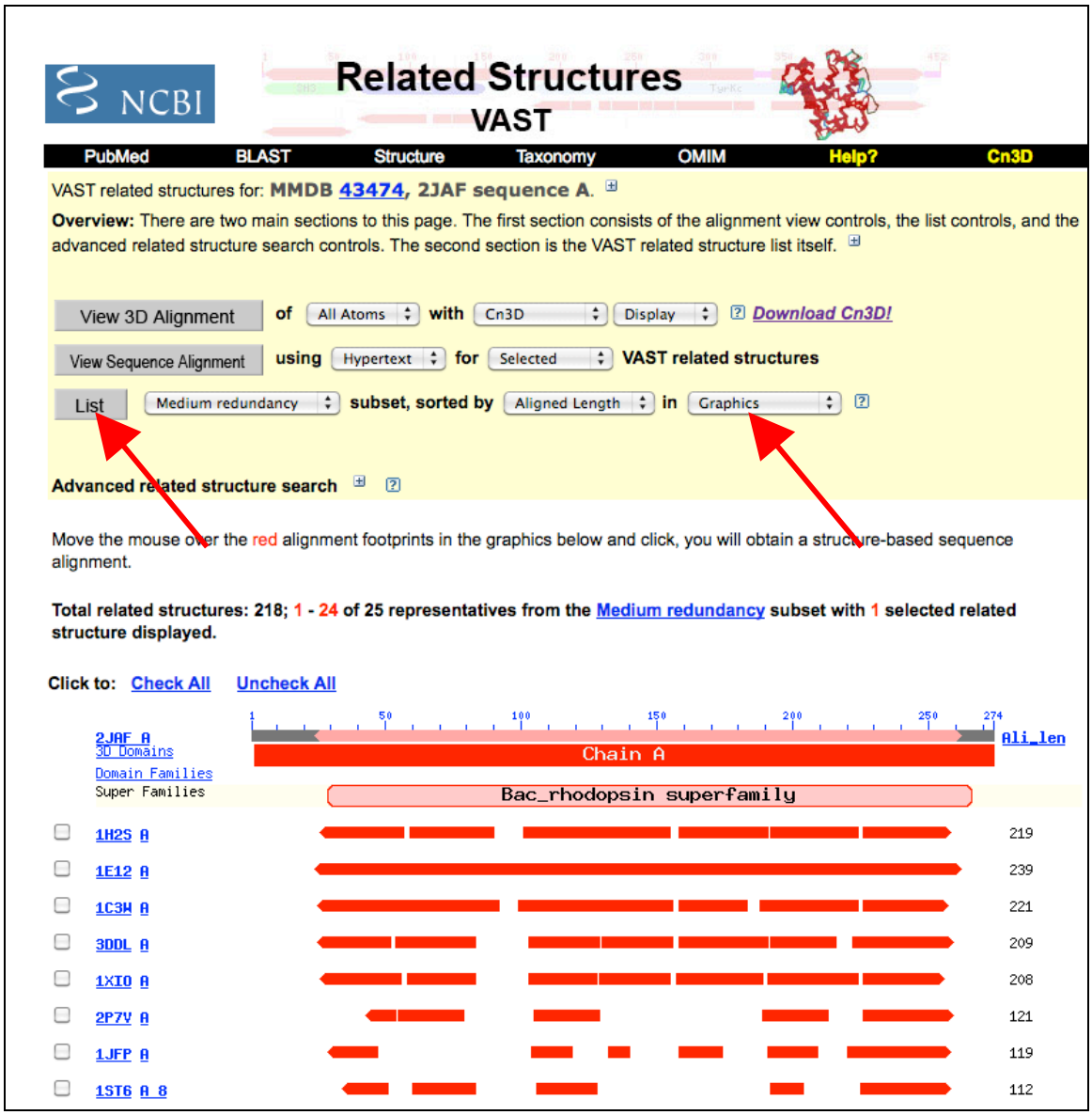

pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin-like protein
The bacterial opsins are retinal-binding proteins that provide light- dependent ion transport and sensory functions to a family of halophilic bacteria. They are integral membrane proteins believed to contain seven transmembrane (TM) domains, the last of which contains the attachment point for retinal (a conserved lysine). This family also includes distanyl related proteins that do not contain the retinal binding lysine and so cannot function as opsins.
Links
Statistics
Structure
Structure View
Program: Cn3D
Drawing: All Atoms
Aligned Rows: up to 5
Download Cn3D
PubMed References
Two pumps, one principle: light-driven ion transport in halobacteria. Trends Biochem. Sci. 1989 Feb; 14(2):57-61
Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor. EMBO J. 1989 Dec 20; 8(13):3963-3971
Electron-crystallographic refinement of the structure of bacteriorhodopsin. J. Mol. Biol. 1996 Jue 14; 259(3):393-421
pfam01036 is a member of the superfamily cl02333.
Sequence Alignment
Reformat Format: Compact Hypertext Row Display: up to 5 Color Bits: 2.0 bit Type Selection: top listed sequences
2JAF_A 29 .[16].LVFVYM.[1].RTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGMIEMP.[11].SQWGRYLTWALSTPMI 119
query 21 .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[5].IYYARYADWLFTTPLL 105
gi 461609 34 .[16].LLFVYM.[1].RNVEDPRAQLIFVATLMVPLVSISSYTGLVSGLTVSFLEMP.[11].TPWGRYLTWALSTPMI 124
gi 2499386 7 .[16].LYFIAR.[1].WGETDSRRQKFYIATILITAIAFVNYLAMALGFGLTIVEFA.[5].IYWARYSDWLFTTPLL 91
gi 2499387 14 .[16].LYFIAR.[1].WSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[5].IYWARYTDWLFTTPLL 98
2JAF_A 120 LLALGLLADVDLGSLFTVIAADIGMCVTGLAAAMT.[3].LLFRWAFYAISCAFFVVVLSALVTDWAASASSA GT 192
query 106 LLDLALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[2].PLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EV 179
gi 461609 125 LIAVGLLAGSNTTKLFTAVVADIGMCVTGLAAALT.[3].YLLRWVWYAISCAFPVVVLYILLAEWAEDAEIA GT 197
gi 2499386 92 LYDLGLLAGADRNTITSLVSLDVLMIGTGLVATLS.[8].GAERLVWWGISTAFLLVLLYFLFSSLSGRVADL.[2].DT 171
gi 2499387 99 LYDLALLAGADRNTIYSLVGLDVLMIGTGALATLS.[8].GAERLVWWGISTGFLLVLLYFLFSNLTDRASEL.[2].DL 178

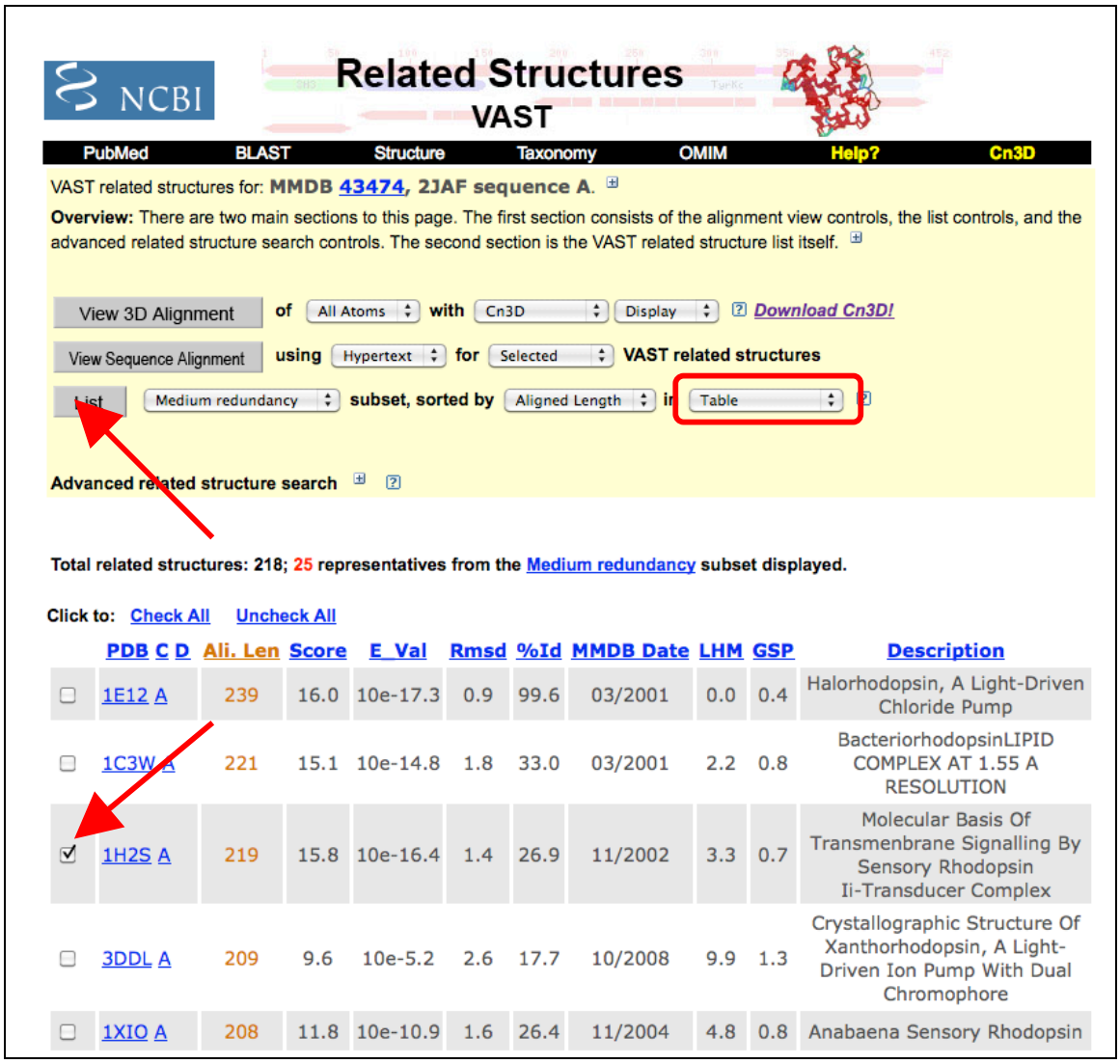

NCBI
Protein
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Format: GenPept FASTA Graphics More Formats▼
Download▼ Save▼ Links▼
PDB: 2JAFA
Chain A, Ground State Of Halorhodopsin T203v
Comment Features Sequence
LOCUS 2JAF_A 274 aa linear BCT 24-SEP-2008
DEFINITION Chain A, Ground State Of Halorhodopsin T203v.
ACCESSION 2JAF_A
VERSION 2JAF_A GI:122920878
DBSOURCE pdb: molecule 2JAF, chain 65, release Aug 27, 2007;
deposition: Nov 28, 2006;
class: Membrane Protein;
source: Mol_id: 1; Organism_scientific: Halobacterium Salinarium;
Strain: R1; Other_details: Dsm 671;
Exp. method: X-Ray Diffraction.
KEYWORDS .
SOURCE Halobacterium salinarum
ORGANISM Halobacterium salinarum
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halobacterium.
REFERENCE 1 (residues 1 to 274)
AUTHORS Gmelin,W., Zeth,K., Efremov,R., Heberle,J., Tittor,J. and
Oesterhelt,D.
TITLE The crystal structure of the L1 intermediate of halorhodopsin at
1.9 angstroms resolution
JOURNAL Photochem. Photobiol. 83 (2), 369-377 (2007)
Change Region Shown
Customize View
Analyze This Sequence
Run BLAST
Identify Conserved Domains
Protein 3D Structure
Ground State Of Halorhodopsin T203v
PDB: 2JAF
Source: Halobacterium salinarum
Method: X-Ray Diffraction
Resolution: 1.7 Å
Identical Proteins for 2JAFA
Chain A, L1-Intermediate Of Halorhodopsi [2JAG_A]
» See all...

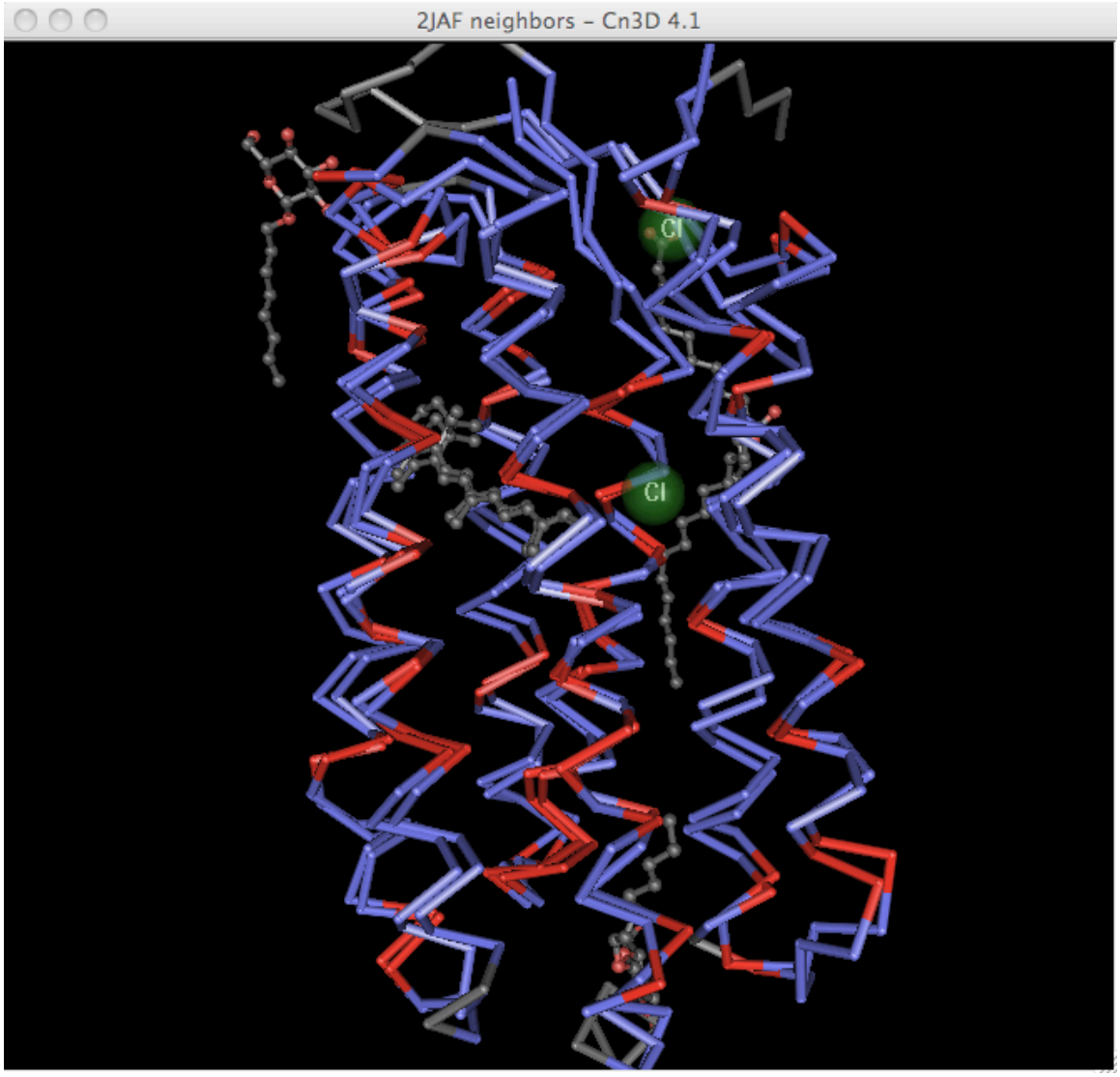

/region_name="Domain 1"
/note="NCBI Domains"
SecStr 27..51
/sec_str_type="helix"
/note="helix 1"
Region 29..266
/region_name="Bac_rhodopsin"
/note="Bacteriorhodopsin-like protein; cl02333"
/db_xref="CDD:154859"
SecStr 58..82
/sec_str_type="helix"
/note="helix 2"
SecStr 85..91
/sec_str_type="sheet"
/note="strand 1"
SecStr 100..106
/sec_str_type="sheet"
/note="strand 2"
Region 110..189
/region_name="Domain 2"
/note="NCBI Domains"
SecStr 112..127
/sec_str_type="helix"
/note="helix 3"
SecStr 131..153
/sec_str_type="helix"
/note="helix 4"
SecStr 159..185
/sec_str_type="helix"
/note="helix 5"
SecStr 193..216
/sec_str_type="helix"
/note="helix 6"
SecStr 227..256
/sec_str_type="helix"
/note="helix 7"
Het bond(242)
/heterogen="(RET, 900)"
ORIGIN
1 msitsvpgvv dagvlgaqsa aavrenalls sslwvnvala giailvfvym grtirpgrpr
61 liwgatlmip lvsissylgl lsgltvgmie mpaghalage mvrsqwgryl twalstpmil
121 lalglladvd lgslftviaa digmcvtgla aamttsallf rwafyaisca ffvvvlsalv
181 tdwaasassa gtaeifdtlr vlvvvlwlgy pivwavgveg lalvqsvgat swaysvldvf
241 akyvfafill rwvannertv avagqtlgtm ssdd
//
LOC100004992 similar to zinc finger protein 484 [Danio rerio]
LOC100004992 [Danio rerio]
» See more...
All links from this record
BLink
Related sequences
Identical proteins
3D domains
Conserved domains
Domain relatives
PubMed
Related structure
Structure
Taxonomy
LinkOut

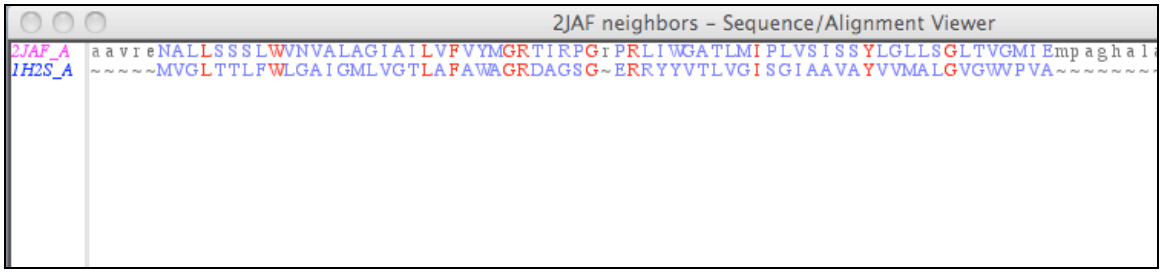

NCBI
Structure
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Structure for Go Clear
Limits Preview/Index History Clipboard Details
Display Summary Show 20 Sort By Send to Download Cn3D
Links: Literature, Domains, Chemicals, Other Links
All: 1 X-ray: 1
1: 2JAF
Related Structures, Literature, Domains, Chemicals, Other Links
Ground State Of Halorhodopsin T203v [Membrane Protein]
Taxonomy: Halobacterium salinarum
Proteins: 1; Chemicals: 4
modified: 2007/10/11; MMDB ID: 43474
Recent activity
Turn Off Clear
Structure Links for Prote... (1)
Structure
Chain A, Ground State Of Halorhodopsin T203v
2JAFA (1)
Protein
...[gi:294946473]
LOC100004992 similar to zinc finger protein 484 [Danio rerio]
» See more...


NCBI
Structure Summary
MMDB
HOME SEARCH SITE MAP
Entrez Structure Protein CDD PubMed Taxonomy PubChem Help Cn3D
MMDB ID: 43474
PDB ID: 2JAF
Search PDB or MMDB ID
Reference: Gmelin W, Zeth K, Efremov R, Heberle J, Tittor J, Oesterhelt D The crystal structure of the L1 intermediate of halorhodopsin at 1.9 angstroms resolution Photochem. Photobiol. v83, p.369-377
The mutant T203V of the light driven chloride pump halorhodopsin from Halobacterium salinarum was crystallized and the X-ray structure was solved at 1.6 angstroms resolution. The T203V structure turned out to be nearly identical to the wild type protein with a root mean square deviation of 0.43 angstroms for the carbon alpha atoms of the protein backbone....
» View full abstract
Description: Ground State Of Halorhodopsin T203v.
Deposition: 2006/11/28
Taxonomy: Halobacterium salinarum
Related Structure: VAST
Structure View in Cn3D
Structure View in RasMol
Tasks: Display Drawing: All Atoms
Download Cn3D
View Cn3D Tutorial
Molecular components in the MMDB structure are listed below and may include macromolecular chains, 3D domains, protein classifications (domain families), and ligands, as available. Mouse over each icon for more information on the component.
Protein
3d Domains
Domain Families
Super Families
Sequence A
1
2
1
Bac_rhodopsin superfamily


NCBI
Related Structures
VAST
PubMed BLAST Structure Taxonomy OMIM Help? Cn3D
VAST related structures for: MMDB 43474, 2JAF sequence A.
Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced related structure search controls. The second section is the VAST related structure list itself.
View 3D Alignment of All Atoms with Cn3D Display Download Cn3D!
View Sequence Alignment using Hypertext for Selected VAST related structures
List Medium redundancy subset, sorted by Aligned Length in Graphics
Advanced related structure search
Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.
Total related structures: 218; 1 - 24 of 25 representatives from the Medium redundancy subset with 1 selected related structure displayed.
Click to: Check All Uncheck All
2JAF A
3D Domains
Domain Families
Super Families
Chain A
Bac_rhodopsin superfamily
Ali_len
1H2S A 219
1E12 A 239
1C3W A 221
3DDL A 209
1XIO A 208
2P7V A 121
1JFP A 119
1ST6 A 8 112


NCBI
Related Structures
VAST
PubMed
BLAST
Structure
Taxonomy
OMIM
Help?
Cn3D
VAST related structures for: MMDB 43474, 2JAF sequence A.
Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced related structure search controls. The second section is the VAST related structure list itself.
View 3D Alignment
of
All Atoms
with
Cn3D
Display
Download Cn3D!
View Sequence Alignment
using
Hypertext
for
Selected
VAST related structures
List
Medium redundancy
subset, sorted by
Aligned Length
in
Table
Advanced related structure search
Total related structures: 218; 25 representatives from the Medium redundancy subset displayed.
Click to: Check All Uncheck All
PDB C D
Ali. Len
Score
E_Val
Rmsd
%Id
MMDB Date
LHM
GSP
Description
1E12 A
239
16.0
10e-17.3
0.9
99.6
03/2001
0.0
0.4
Halorhodopsin, A Light-Driven Chloride Pump
1C3W A
221
15.1
10e-14.8
1.8
33.0
03/2001
2.2
0.8
BacteriorhodopsinLIPID COMPLEX AT 1.55 A RESOLUTION
1H2S A
219
15.8
10e-16.4
1.4
26.9
11/2002
3.3
0.7
Molecular Basis Of Transmenbrane Signalling By Sensory Rhodopsin Ii-Transducer Complex
3DDL A
209
9.6
10e-5.2
2.6
17.7
10/2008
9.9
1.3
Crystallographic Structure Of Xanthorhodopsin, A Light-Driven Ion Pump With Dual Chromophore
1XIO A
208
11.8
10e-10.9
1.6
26.4
11/2004
4.8
0.8
Anabaena Sensory Rhodopsin


2JAF neighbors – Cn3D 4.1
Cl
Cl


2JAF neighbors – Sequence/Alignment Viewer
2JAF_A aavreNALLSSSLWVNVALAGIAILVFVYMGRTIRPGrPRLIWGATLMIPLVSISSYLGLLSGLTVGMIEmpaghal
1H2S_A ~~~~~MVGLTTLFWLGAIGMLVGTLAFAWAGRDAGSG~ERRYYVTLVGISGIAAVAYVVMALGVGWVPVA~~~~~~~~

Problem 2

In this problem, we will follow these steps:

A. Identify conserved domain(s) present in a protein.
B. Search for other proteins containing similar domain(s).
C. Explore a 3D modeling template for the query sequence.
D. Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

CDD

Paste the following protein sequence in the CD-Search query box and run the search.

MSATKKTYSSTTSAKSKHSVRVAQTTADAALEAVYEMSGDSGDSFDYSKSVG
QSAESVPAGAVTAYLQRMQREGLIQNFGCMVAVEEPNFCVIAYSENASEFLDLI
PQAVPSMGEMDVLGIGTDIRTLFTPSSSAALEKAAATQDISLLNPITVHCRRSGK
PLYAIAHRIDIGIVIDFEAVKMIDVPVSAAAGALQSHKLAARAITRLQALPGGDIEL
LCDTIVEEVRELTGYDRVMAFKFHEDEHGEVVAEIRRMDLEPYMGLHYPATDIP
QASRFLLMKNRVRLIADCYASPVKLIQDPDIRQPVSLAGSTLRAPHGCHAQYMG
NMGSIASLVMAVIINDNEEYSRGAIQRGRKLWGLVVCQHTSPRTVPFPLRSVCE
FLMQVFGMQLNLHVELAAQLREKHILRTQTLLCDMLLRDAPIGIVSQTPNIMDLV
KCDGAALYYGKRVWLLGTTPTENQIKEIADWLLEHHNDSTGLSTDSLADANYP
GAHLLGDAVCGMAAAKITAKDFLFWFRSHTATEVKWGGAKHDPDEKDDGRKM
HPRSSFKAFLEVVNKRSPPWEDVEMDAIHSLQLILRGSFRDIADSDTKTMIHARL
NDLKLQGVEERNALANEMSRVLETAAAPILAVDSRGMINAWNAKIAQVTGLPVE
EAMHCSLTKDLVLDESVVVVERLLSLALQGEEEQNVEIKLKTFGTQTTERAVILIV
NACCSRDASDFVVGVFFVGQDVTEQRMFMDRFTRIQGGEKTTVQDPHPLMRP
SFDGDEFGRTFKRNSALGGLKDHATGSVERLDLYLRRAEECMEVMETIPSPKF
NNKQCQYLAGKLKAVLQSASLFLRISHHEHHELGASIDMGRHVEIFKLLLALAKE
IESFIQGCCKDEWIKAAMTLTNVSEYVSSMGFNLELCKIAFCKSCAASGSLTLDQ
IEVICKDEAEVVKRNASIDVDTLFAKVIYDLTEKTLSSDQNDLAIYLLQRLKRAKPI
LPSFSSRPSWWNFYDDWSFSEKFFQWIQITGSLGSGSSATVEKAVWLGTPVA
KKTFYGRNNEDFKREVEILAELCHPNITSMFCSPLYRRKCSIIMELMDGDLLALM
QRRLDRNEDHDSPPFSILEVVDIILQTSEGMNYLHEKGIIHRDLKSMNILVKSVKV
TKSEIGYVHVKVADFGLSKTKDSSTRYSNQTWNRGTNRWMAPEVINLGYESTE
GEISFDGKVPKYPLKSDVYSFGMVCYEVLTGDVPFPEEKNPNNVKRMVLEGVR
PDLPAHCPIELKALITDCWNQDPLKRPSFAVICQKLKYLKYLLMTGFSSYQDSYP
STEEPS

A. What are the domains present in this protein?

-Suppose, we are interested in the serine/threonine protein kinase domain.

Obtain more information about it by searching in NCBI's Bookshelf

B. Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domains architectures" button. To display the records, click on the links to the subsets of sequences and from there on the “Look up Sequences in Entrez”. Change the display from “Summary” to “FASTA”.

C. Go back to the CD-Search results page. Click on the “Full Report” radio button. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic representation of the serine/threonine kinase domain from CDD (CDD|00180, the second pink graphic link from the top). Use the "Aligned Rows" list box pull down menu to specify "up to 5" sequences and invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Structure View" button.

To show only one top structure, click on the down arrow key. For better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a serine/threonine protein kinase active-site signature (IIHRDLKSMNILV) where K is the ATP binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the first protein sequence.

Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the “Done” button. To zoom in, press z on the keyboard. Note the heterogen near the conserved lysine residue.

D. To obtain the structural neighbors for the serine/threonine protein kinase protein, first click on the top structure entry link 1F3M_C of the similar protein from the CD-Browser page. Next, click on the structure link on the bottom right side, then on the 1F3M header for the record, and finally on the grey “chain C” graphic. Select one or more of the check boxes next to the structure neighbors and download the structures by clicking on the "View 3D Alignment" button.